STRUCTURAL ANALYSES OF TELOMERE ASSOCIATED PROTEINS

by

Feng Wang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biological Chemistry)
in The University of Michigan
2010

Doctoral Committee:

Assistant Professor Ming Lei, Chair
Professor Robert S. Fuller
Professor Janet L. Smith
Associate Professor Gary D. Hammer
Associate Professor Zhaohui Xu

2010

This dissertation is dedicated to my parents: Xiaobao Wang and Yanchun Liu. Without their endless love and support, I would not be able to achieve what I have accomplished.

Acknowledgements

First of all, I would like to thank my mentor Dr. Ming Lei. During my days in the Lei lab, Ming taught me everything he knows hand by hand. Besides all the details in each experiment, Ming shared me with all his valuable experience he gained during his own career. Ming taught me how to think as a scientist, how to make presentations, how to write scientific assays… there are so many that I can't list them all. I have learnt so much from Ming, not only the knowledge in science, but also the attitude towards life and people. Without his directions in science and help in life, I could not pursue this far. I thank Ming from the bottom of my heart.

I thank all the members of my dissertation committee: Dr. Robert S. Fuller, Dr. Janet L. Smith, Dr. Gary D. Hammer and Dr. Zhaohui Xu, for their valuable directions and suggestions at every stage in my Ph.D study.

I would like to thank all the members in Lei lab. I want to thank Dr. Yuting Yang for her great help with the crystallography in each of my project. I want to thank Dr. Yong Chen for offering me valuable suggestions and discussions during my research. I want to thank Paul Baciu and Ke Wan for their help in my POT1-TPP1 project and BLM complex project. I also thank Dr. Zhixiong Zeng, Jia Sun, Dr. Wei Deng, Laura A. Confer, Dr. Bingbing Wang and Dr. Jing Huang. All the members in the Lei lab are so nice and I feel so lucky to work with them.

I want to thank all my collaborators: Dr. Thomas Cech, Dr. Elaine Podell, and Dr. Arthur Zaug in the POT1-TPP1 project, Dr. Neal Lue, Dr. Eun Young Yu and Dr. Erik Zuiderweg in the Est3 project, Dr. Julia Cooper and Dr. Piere-Marie Dehe, Dr. Hongyu Hu and Ziren Zhou in the Taz1 project, Dr. Patrick Sung and Dr. Valeria Busygina in the RMI1-RMI2 project.

At last but not least, I want to thank all the faculty members, staffs and students in the Department of Biological Chemistry, University of Michigan. Especially I want to thank Elizabeth L. Goodwin, for her assistance for any of my questions during my graduate study.

Table of Contents

List of Figures

Figure

List of Tables

Table

ABSTRACT

Telomeres are specialized protein-DNA complexes that compose the natural termini of linear chromosomes. Telomeres function as chromosome caps and prevent chromosome ends from undergoing deleterious degradation and fusion events. In most eukaryotes, telomere length is maintained by telomerase, a specialized reverse transcriptase, which adds telomeric DNA to the 3' ends of chromosomes to ensure complete genome replication. In this work, I studied the structures and functions of three different telomere associated protein complexes.

POT1 binds the single-stranded telomere overhangs at human chromosome ends and suppresses unwanted DNA repair activities. TPP1 is a binding partner of POT1 that forms part of a six-protein shelterin complex at telomeres. The crystal structure of TPP1 reveals a structural similarity to the β-subunit of the telomere end-binding protein of a ciliated protozoan, suggesting that TPP1 is the missing β-subunit of human POT1 protein. In addition, structural and bioinformatic analyses suggest that TPP1 is evolutionarily conserved and the yeast telomerase component Est3 is an ortholog of TPP1. Telomeric DNA end-binding proteins have generally been found to inhibit, rather than stimulate, the action of telomerase. In contrast, we find that TPP1 and POT1 form a complex that increases the activity and processivity of human telomerase. We propose

that the POT1–TPP1 complex is a processivity factor for telomerase during telomere extension.

Taz1, the fission yeast double-stranded telomeric DNA binding protein, has long been considered as the structural homologue of mammalian shelterin proteins TRF1 and TRF2. Both TRF1 and TRF2 contain a central TRFH domain, and the dimerization of this domain is crucial for telomere localization of the TRF proteins. However, because of low sequence identity between Taz1 and the TRF proteins, a definitive answer to their structural similarity is still unknown. Here, my crystal structure of the Taz1 TRFH domain shows that although $Taz1_{TRFH}$ is still an α-helical structure, it is not a structural homolog of the TRFH domains of human TRF1 and TRF2. Notably, $Taz1_{TRFH}$ is a monomer, and the dimerization of Taz1 is mediated by another domain outside of TRFH. The structure of the Taz1 dimerization domain reveals that Taz1 employs a different architectural principle for homodimerization. In addition, we also determined the NMR structure of Taz1 complexed with another telomere protein Rap1. Strikingly, with no apparent sequence similarity, the Taz1-Rap1 interaction closely resembles that of human TRF2-RAP1.

The BLM-TOPOIIIα-RMI1/RMI2 (BTR) complex is a DNA structure-specific "dissolvasome". BTR is essential for genome integrity and potentially plays an important role in telomere maintenance in telomerase-negative cancer cells. However, little is known about the molecular architecture of the BTR complex and how this multi-subunit machinery dissolves homologues recombination intermediates such as D-loops and double Holliday Junctions. To address these questions, I carried out structural studies of RMI1 and RMI2, the two regulatory subunits of the BTR complex. My data demonstrate

that both RMI1 and RMI2 are OB-fold containing proteins. RMI2 binds to the C-terminus of RMI1 to form a stable RPA-like complex, and this interaction is essential for the stability of the BTR complex. The crystal structure of the N-terminus of RMI1 reveals a non-canonical OB-fold with two extra structural elements that are essential for the “dissolvasome” activity of BTR.

Chapter 1

INTRODUCTION

1.1 An early history of telomere biology

DNA in eukaryotic cells is associated with various proteins to form compact structures called chromosomes, which carry the genetic information. During cell division, when the sister chromosomes are replicated and separated, the genetic information is copied and passed down from mother cells to daughter cells. Therefore, the stability and integrity of the chromosomes are crucial to living organisms. Since the observation of chromosomes and cytokinesis, geneticists and cytogeneticists, including Hermann Muller (1890 – 1967, Nobel laureate 1946) and Barbara McClintock (1902 – 1992, Nobel laureate 1983), had contributed a lot in understanding the nature of chromosomes. In the 1920s, by treating fruit flies with X-rays, Muller observed several kinds of chromosome breaks such as inversions, translocations and deficiencies, and he was able to recover them by using his new genetic techniques. However, Muller failed to recover the chromosome terminal deficiencies. His explanation was that the recovered chromosomes were usually the result of the rejoining of two broken ends. However, such rejoining could not occur between "originally free ends" or between broken ends and "originally free ends". Muller realized that the chromosome ends might have a more important function than people ever imagined. In 1938, Muller, his colleague Darlington and independently, J.D.S. Haldane

named the “free end” the “telomere” (The remaking of chromosomes. *The Collecting Net*). But at that time, Muller did not have a clear idea about the nature of telomeres. He hypothesized that each telomere contains a gene that cells could not lose.

At the same time when Muller studied the defects in fruit flies caused by X-rays, Barbara McClintock was pursuing genetics and developmental research using another powerful system, maize. McClintock developed new microscopes that allowed her to visualize individual maize chromosomes. McClintock observed special chromosome “knobs” (locally heterochromatic regions), some of which were located at the chromosome ends. She recognized that these “knobs” at the chromosome ends might be very important and named them as the “natural ends” of the chromosomes (McClintock B. 1931. Cytological observations of deficiencies involving known genes, translocations and an inversion in *Zea mays*.). In agreement with Muller, McClintock pointed out that the intact chromosome ends have a unique function that is different from broken chromosome ends caused by X-ray irradiation, as the broken ends never fused with the “natural ends”. More importantly, McClintock also noticed that the broken ends could be healed and “remain permanently healed” in some early embryo tissues [1]. Based on her discoveries, McClintock reasoned that there must exist some mechanism that could heal a single broken end “during the reproductive cycle of the chromosome” (Maize genetics. 1942-1943. *Carnegie Inst. Wash. Year Book*). McClintock’s insight perfectly pioneered the modern understanding that telomerase is actively involved in healing broken ends during S-phase. These pioneering discoveries on the “natural chromosome ends” by Muller and McClintock laid out the foundations of modern telomere biology.

1.2 The Hayflick limit

Besides maintaining the integrity of the chromosomes, telomeres are also related to cellular senescence. In the early 1960s, Leonard Hayflick at the University of Pennsylvania first observed that a population of normal human fetal cells in cell culture could only divide between 40 and 60 times (known as the Hayflick limit) before the cells entered the senescence phase [2-4]. In 1974, Alexei Olovnikov proposed that the non-coding telomeric nucleotide repeats might provide a buffer region for the chromosomal shortening during DNA replication [5]. He proposed that the number of the telomeric repeats (or the telomere length) determines how many rounds the cells could divide. It was not until more than twenty years later that Cooke and Smith found convincing evidence to support Olovnikov's proposal. In 1986, Cooke and Smith measured the average telomere length difference between the sex chromosomes from sperm cells and those from adult cells. The sperm cells were found to have longer telomere length than the adult cells [6]. It was just at that time that telomerase was discovered in *Tetrahymena* (see Section 1.5) [7]. Cooke and Smith reasoned that the telomere length and telomerase regulation could be linked with cell aging. Later observations confirmed this hypothesis.

1.3 The end replication problem

In the 1950s and 1960s, it had been revealed that all eukaryotic chromosomes are made of linear DNA molecules. The replication of the double-stranded DNA is semi-conservative and each strand of the double helix is used as the template for the new stand synthesis [8]. The newly synthesized DNA strand that is synthesized in the 5' to 3' direction is defined as the leading stand (Fig. 1.1). On the opposite side is the lagging

strand, which runs in the 3’ to 5’ direction. Both the leading and the lagging strands are synthesized by DNA polymerase. During the replication process, initiated by a short RNA primer synthesized by primase, the DNA polymerase can only go in the direction from 5’ end to 3’ end. Therefore the leading strand can be synthesized continuously while the lagging strand is synthesized in short fragments that are referred to as Okazaki fragments [9, 10]. However, at the end of each round of DNA replication, when the last RNA primer was removed, the terminal DNA cannot be synthesized by polymerase, leaving a gap that is up to 500 nucleotides (mean value of 250 nucleotides) un-replicated at the 5' end of the chromosome [11, 12] (Berg JM, Tymoczko JL, Stryer L, Clarke ND. 2002. *Biochemistry* Chapter 27: DNA Replication, Recombination, and Repair). This is the famous “end replication problem” initially raised by James Watson in 1972 [13] (Fig. 1.1). If the cells continue to lose terminal sequences, they will stop growing and undergo cellular senescence. Therefore, the “end replication problem” is one important reason why cultured cells are only able to divide a certain numbers of generations [3] (also known as the Hayflick limit).

The “end replication problem” would have to be solved by telomeres. Before the molecular details about telomeres or their DNAs were revealed, some knowledge had been gathered about the replication properties at the telomeres. Virus and bacteria bearing temporarily or permanently circular genomes won’t be bothered by the “end replication problem”. T4 and T7 virus that have linear genomes can overcome the problem by genome ends fusions so that they are able to use the adjacent DNA strand as the primer [13, 14]. For some species such as the mammalian adenoviruses, the Terminal Protein (TP) acts directly as the primer for initiation of replication of the linear genome [15-19].

However, a puzzle still remained about how eukaryotic double-stranded linear DNA is replicated at the telomeres. One proposed solution was that of Cavalier-Smith (1974), who suggested that the sequences at the ends of eukaryotic linear chromosomes could be palindromic. Following DNA replication, the very end that carries the 3'-OH group would be able to fold back and pair with itself to form a terminal loop (Fig. 1.2). The remaining gap would be closed by the DNA polymerase fill-in and ligation. Then, the opposite strand could be nicked by a specific endonuclease to make a new 3'-OH end for priming after the loop had unfolded to become the template for the last step to fill in the gap by DNA polymerase [20]. Although Cavalier-Smith's model was proved wrong, it brought insights in understanding the telomere replication machinery.

1.4 Telomeric DNA sequences

The identification of telomeric DNA sequences is essential for understanding telomere replication and chromosome end structure. However, this was not an easy task due to the fact that chromosomes are too long for direct sequencing. The limited number of chromosomes in somatic cell also made it difficult to gather enough telomere material. The discovery of minichromosomes that contain amplified ribosomal RNA genes (rDNAs) in some simple eukaryotes, such as the ciliated protozoan *Tetrahymena*, made it possible to sequence the telomeric DNA [21, 22]. Compared with regular chromosomes, these linear DNA minichromosomes are relatively shorter and abundant, providing enough telomere materials for the analyzing purpose. In 1976, Engberg, Karrer and Gall first showed that the linear rDNA minichromosomes in cillated protozoan *Tetrahymena thermophila* are palindromic [23, 24]. Later, Blackburn and Gall determined the end

sequence by *in vitro* labeling combined with restriction endonuclease digestion, depurination and fingerprinting analyses. The digested fragments were very heterogeneous in length and contained tandem hexanucleotide sequence repeats 5'-(CCCCAA)$_n$-3' (*n* is between 20 and 70) [25]. In the 1980s, Szostak and Blackburn designed a linear plasmid that allowed them to clone out the budding yeast (*Saccharomyces cerevisiae*) telomeres and mapped the sequences [26]. Soon after, telomeric DNA sequences from many other eukaryotes were determined. Generally, the telomeric DNA is made of tandem repeated sequences, which are usually guanine-riched in the 5' - 3' strand (also called G-strand). Most of the telomeric DNA sequence is conserved among all the chromosomes in one species, and is usually different from species to species. The telomeric DNA sequence in different species is summarized in TABLE 1.1. In vertebrates, the telomeric repeat sequence is "TTAGGG". Notably, the DNA end is not blunt; there is a 3' overhang of the G-riched strand extending at the DNA termini [27-29].

1.5 A specialized reverse transcriptase, telomerase that synthesizes telomeric DNA

In the semi-conservative DNA replication process, the newly synthesized strands cannot be fully replicated from the templates due to the end replication problem (Section 1.2), which causes telomere shortening in each round of DNA replication. It was predicted that the average rate of the telomere erosion in human cells would be about 63 base pairs in each cell division [30, 31]. If there were no mechanism to compensate this telomere attrition, chromosome ends would eventually lose the protection by telomeres, and the "nude" chromosome ends would be recognized as DNA damage sites. In consequence,

the cells would trigger downstream DNA repair pathways. Inappropriate chromosomal end-to-end fusions through non-homologous end joining (NHEJ) or homologous recombination (HR) pathways would create dicentric chromosomes, which are unstable in mitosis (and meiosis) and lead to genome instability [32-35]. Therefore, all eukaryotic organisms have evolved mechanisms to add telomere sequences to the chromosome ends and thus counterbalance the terminal DNA attrition caused by the "end replication problem". Several hypotheses, including the "hairpin" model and the recombination model, had been proposed to solve the end replication problem. However, in yeast cells, telomeric repeats could be added to a plasmid end that is capped with *Tetrahymena* telomeric repeats, suggesting that telomere addition doesn't require DNA templates [36]. Therefore, it was summarized that there must be an unknown enzyme that is responsible for the de novo telomere repeats addition to the chromosome ends.

In 1985, Carol Greider and Elizabeth Blackburn first successfully demonstrated the activity of this enzyme from the extracts of mating *Tetrahymena*. This enzyme, known as telomerase, could add the correct *Tetrahymena* telomeric repeats TTGGGG to the 3' end of an oligonucleotide [7]. More importantly, telomerase has the ability to recognize the ending sequence of the oligonucleotide that it is elongating. For example, if the oligonucleotide ends with TTG, telomerase will first add GGG and then start the next round of TTGGGG addition. In other words, the sequence at the 3' end determines what telomerase will first add to the primer [37]. This special feature implied that telomerase might use an internal nucleic acid template to mediate the nucleotide addition. This was later confirmed by the isolation of telomerase RNA, which was used by telomerase as the replication template [38]. The RNA template of *Tetrahymena* telomerase contains one

and a half telomeric repeats (5'-AACCCCAA-3') and helps telomerase discern the end sequence [38]. When the RNA template was mutated, telomerase was found to add mutated telomeric repeats to the end of the primer [39]. This confirmed that telomerase is a ribonucleoprotein (RNP) reverse transcriptase. The discovery of telomerase solved Watson's end replication problem. Since then, telomerase was identified in many other organisms. In 1996, Lungblad and coworkers identified three yeast telomerase components (*EST1*, *EST2* and *EST3*) through genetic screening in *Saccharomyces serevisiae* [40]. At the same time, telomerase was also purified from the hypotrichous ciliate *Euplotes aediculatus* by using an anti-sense oligonucleotide bait that was complimentary to the telomerase RNA template by Cech and coworkers [41].

Human telomerase includes an RNA template (telomerase RNA component, TERC, also known as TR or TER) [38, 42-45], a highly conserved reverse transcriptase catalytic subunit (telomerase reverse transcriptase, TERT) and some other accessory factors (for instance, Est1 and dyskerin in humans) [32, 45-50]. The RNA hTR and the catalytic subunit hTERT form the core of human telomerase. Two different genes in the human genome code for TERC and hTERT separately. TERT contain a C-terminal reverse transcriptase (RT) domain that is similar to RTs from retroelements and retroviruses. There are seven motifs (1, 2, A, B, C, D and E) that can be recognized in the conserved RT domain in all the TERTs (Fig. 1.3a). Three-dimension structural modeling results suggested that similar to the RT domains from human immunodeficiency virus (HIV) and murine leukemia virus (MLV), the seven RT motifs from TERT form a "right hand"-like structure. RT motifs 1 - B are the fingers, motifs C – E form the palm, and the carboxy-terminal extension (CTE) domain forms the thumb (Fig. 1.3b) [51-55]. The

active site of the enzyme is found within motifs A – C. Three aspartic acids from A and C are most important for composing the active site, as shown in the crystal structure of *Tribolium castaneum* TERT (Asp251, Asp343 and Asp344) [56]. The hTR gene, encodes an untranslated RNA which is 451 nucleotides in length that remains as an RNA in functional telomerase. The template region of human TERC is 3'-CAAUCCCAAUC-5', which is complementary to the human telomeric DNA sequence TTAGGG. Human TERC can be divided into four regions, the template region, the TERT binding site, the pseudoknot domain, and the binding region for associated proteins such as the Dyskerin complex (Gar1, Nop10, Nhp2, Dyskerin) (Fig. 1.3c) [57-61]. Notably, the RNA template regions from different species share several common features. First, the RNA template is single-stranded and about 1.5 – 2 times of the telomeric repeats in length. The length of the template allows TERT to accomplish annealing and nucleotides addition within the same elongation step. Second, when the telomerase holoenzyme is assembled, the template will be buried into the TERT active site. TR and TERT are the most essential telomerase components and telomerase activity can be reconstituted *in vitro* when mixing TR and TERT properly under certain conditions. In my studies, HA-tagged hTERT and hTR were *in vitro* T7-transcribed/translated in the rabbit reticulocyte lysate and then the reconstituted enzyme was affinity purified by anti-HA conjugated Sepharose. Robust telomerase activity could be detected in direct telomerase activity assays by using the *in vitro* reconstituted telomerase via this method [62, 63].

After telomerase holoenzyme is properly assembled, TERT forms a special "mitten" structure to warp the chromosome end in order to favor the telomeric repeats addition [45]. When telomerase adds nucleotides to telomere ends, it takes the following

steps: annealing, elongation, translocation and further elongation in a processive manner [64, 65]. Firstly, with the help of the accessory factors, telomerase locates to the chromosome end and anneals to the 3’ G-overhang) via its internal RNA template (Fig. 1.4a). Then the protein component TERT functions as the reverse transcriptase and adds nucleotides to the end of the G-overhang according to the RNA template (Fig. 1.4b). After the initial round of nucleotide extension, the RNA template/telomere DNA hybrid duplex is disassembled and the telomere end realigns to the 3’ end region of the template (the “Translocation” step to the end “TTAG” sequence in this case, as shown in Fig. 1.4c) before starting the next round of “GGTTAG” nucleotide addition (the “Elongation” step as shown in Fig. 1.4d). Telomerase repeats the translocation and elongation steps, and add “GGTTAG” repeats to telomeres continuously (Fig 1.4e-f). Telomerase adds telomeric repeats to telomere without falling off the DNA and it is called repeat addition processivity of telomerase, which is defined as “the number of the bases synthesized when the cumulative probability of dissociation is ½” [64]. Specifically, Processivity = $R_{1/2}$ = -ln2/(2.303k), where k is the slope and $R_{1/2}$ is the number of repeats synthesized before half of the chains have dissociated, analogous to $t_{1/2}$ in radioactive decay [63]. After the telomerase extension at the 3’ G-overhang, the proper processing of 5’ C-strand is also required. The 3’ G-overhang could end with any nucleotide within the “TTAGGG” sequence [66]. However, when telomerase is present, TAG-3’ is most likely to be observed. Unlike the 3’ end, in contrast, the human 5’-strand usually ends with ATC-5’ and some evidence suggests that the C-strand sequence is determined by the single-stranded telomeric DNA binding protein POT1 [67]. However, the mechanism of C-strand processing is still unknown.

1.6 Telomere binding proteins

Telomere binding proteins recognize the short tandem telomere repeats and are essential for stable chromosome maintenance as reflected in the precipitous genome-destabilizing outcomes following gene deletion and over-expression of mutant alleles [32]. The nucleoprotein complexes formed upon association of telomere binding proteins with telomeric DNA distinguish natural chromosome ends from double-stranded breaks and thereby protect chromosome termini from inappropriate end-to-end fusion. Additionally telomere binding proteins recruit and regulate telomerase to ensure an appropriate length of structural DNA that is maintained as a buffer against loss of genetic information stored in genes close to the ends of linear chromosomes. Telomere binding proteins can be broadly divided into two classes on the basis of double-stranded versus single-stranded DNA-binding specificity. DNA-binding specificity correlates with folding motifs particular to each class, with the OB-fold dictating the single-stranded G-overhang recognition and the Myb motif proteins directed towards double-stranded telomere DNA.

1.6.1 Telomeric proteins that recognize the single-stranded G-overhang

The 3’ end single-stranded G-overhang that extends beyond the duplex region of the telomere is a conserved structure in most eukaryotic organisms from ciliated protozoa, to yeasts and mammals [68-70]. The first protein to be identified that specifically recognizes and caps the single-stranded G-overhang was the ciliate *Oxytricha nova* TEBP (telomere end binding protein) [71, 72]. TEBP is composed of two subunits, α and β. This telomere-specific structural protein complex recognizes and binds tenaciously to the

single-stranded G-overhang in a sequence-specific fashion, also protecting the neighboring duplex telomeric DNA [73].

In budding yeast *Saccharomyces cerevisiae*, the single-stranded G-overhang is bound specifically by Cdc13 [74, 75]. *Cdc13* was initially identified as a cell cycle related gene. Loss of *cdc13* led a *rad9*-dependent cell cycle arrest and induced chromosomes ends recombination, indicating a connection between Cdc13 and telomeres [76, 77]. Cdc13 was also identified through a screening for the telomere length defect genes (referred as to ever shorter telomeres) [40, 75]. Cdc13 has two separate functions: it is involved in both chromosome end protection and the recruitment of the telomerase and hence in telomere replication [40, 74].

Due to the very limited amount of materials, initial biochemical studies of single-stranded G-overhang binding proteins failed to deliver the desired factors in organisms other than ciliates. Eventually, POT1 (protecting of telomeres 1), a conserved protein found in fission yeast *Schizosaccharomyces pombe*, humans and many other organisms, was identified through a database search based on a weak sequence similarity to the amino-terminal region of the *O. nova* TEBP α subunit [78-83]. Deletion of the *pot1* gene in fission yeast leads to rapid loss of telomeric DNA and chromosome circularization, suggesting POT1 has a crucial role in telomere capping [78]. POT1 binds to the G-strand of *S. pombe* telomeric DNA with high affinity in a highly cooperative manner [78, 84]. In human cells, POT1 localizes to telomeres, and loss of the G-rich overhang causes a reduction in hPOT1 binding [79, 85]. In human cells, POT1 has a crucial role in telomere length homeostasis, acting as the terminal transducer of telomere length control. Overexpression of a mutant form of human POT1 missing the N-terminal OB fold

(POT1$^{\Delta OB}$) led to rapid telomere lengthening in telomerase-positive cells [85]. In other studies, however, expression of full-length POT1 caused substantial telomere elongation by telomerase [86, 87]. These disparate conclusions indicate that human POT1 plays an essential but complicated role in the regulation of telomere length. Human POT1 also contributes to the protection of chromosome ends, since partial knockdown of POT1 with RNAi results in a DNA-damage response at telomeres, reduction in the single-stranded telomeric DNA, changes in the 5' end of the chromosome, and a telomere fusion phenotype [67, 87, 88].

Mechanistic insights into how the single-stranded G-overhang binding proteins bind and protect telomeres started to emerge after the determination of the three-dimensional structures of the telomere end-binding proteins: TEBP, Cdc13, and POT1 [89, 90] [91-93]. All these proteins bind the single-stranded G-overhang using a conserved DNA-binding motif, the OB (oligonucleotide/oligosaccharide binding)-fold [94]. The OB-fold is a structural domain of 70 – 180 amino acids in length with diverse functions, and has been found in many proteins including human replication protein A (RPA), the B subunit of *E. coli* verotoxin-1 and *E. coli* single-stranded DNA-binding protein (SSB) [94]. OB-folds are difficult to recognize by amino acid sequence analysis alone because of the very low sequence identity, and consequently structural information is often needed for OB-fold identification. The OB-fold comprises two orthogonally packed anti-parallel β sheets with β1: β4: β5 strand topology in one sheet and β1: β2: β3 topology in the other. The N-terminal strand β1 continues as the outer edge of both sheets. Strands β4 and β5 often fold over to extend the other sheet and thus complete a closed β-barrel-like structure. The segment joining strands β3 and β4 often includes a

short helix (Fig. 1.5)[94]. Most telomere-binding OB-folds are further characterized by a C-terminal α-helix. The loops connecting β strands of the OB-fold are variable in length and these insertions/deletions account for the unreliability of current bioinformatics tools for positively identifying OB-folds.

The first well-characterized structure of the single-stranded G-overhang binding protein is *O. nova* TEBP [89]. The crystal structure of the heterotrimeric TEBPα-TEBPβ-ssDNA complex revealed a total of four OB folds. TEBPα consists of three OB folds, two tandem OB folds at the N-terminus (OB1 and OB2) that bind DNA and a third one at the C-terminus (OB3). The single OB-fold in TEBPβ also interacts with DNA [89]. Together, the three DNA-binding OB folds form a deep DNA-binding cleft and the 3'-terminal nucleotide is buried deep within the complex, making it inaccessible to telomerase, consistent with the finding that the TEBPα-TEBPβ complex blocks telomerase activity [95]. The interaction responsible for the TEBPα-TEBPβ association is established by an extended peptide loop contributed by TEBPβ that wraps around the C-terminal OB of the α subunit.

Lei *et al.* solved the crystal structure of the N-terminal part of *S. pombe* POT1 complexed with single-stranded telomeric repeat GGTTAC [92]. The structure confirmed that the OB-fold at the N-terminus of POT1 mediates the specific binding to ssDNA with high affinity. By the two protruding loops at one side of the β-barrel, the POT1 OB-fold clamps the single-stranded fission yeast telomeric DNA repeat "GGTTAC" tightly. The ssDNA was compactly folded and was buried in a basic pocket in the direction from 5' to 3' across the OB-fold [92]. Later on, the crystal structure of the DNA-binding domain of human POT1 reveals two tandem OB folds bound the minimum binding sequence

(TTAGGGTTAG) that is more than one telomere repeat [93]. The two OB folds pack together and form a continuous DNA binding cleft. Notably, the last nucleotide G10 is deeply buried in the binding pocket. It was hypothesized that the burying of G10 may prevent access by telomerase [93]. The superposition of the crystal structures of *O. nova* TEBPα, *S. pombe* POT1 and human POT1shows that the three single-stranded G-overhang-binding proteins bind to their cognate single-stranded telomere DNAs in a similar manner. Taken together, both functional and structural studies strongly support the notion that POT1 proteins are indeed the homologue proteins of *O. nova* TEBP α subunit as initial sequence analysis suggested. Whether higher eukaryotes have a TEBP β subunit homologue, however, remains unknown.

TPP1 is the most recently identified component of human telomere proteins that simultaneously interacts with both POT1 and another human telomere protein TIN2 [96-98]. A 60-residue fragment at the C-terminus of TPP1 binds to the N-terminal half of TIN2, and a central region of TPP1 with ~ 100 amino acids binds to the C-terminal half of POT1. Database searches failed to identify any notable structure features in TPP1. Functional analyses suggested that TPP1 regulates the telomere recruitment of POT1 and TPP1 itself is a negative regulator telomere length [96-98] (reviewed in [99]). In order to explore the additional functions of TPP1 at the telomere, I determined the crystal structure of a portion of TPP1. The structure of TPP1 reveals an OB fold that is structurally most similar to the β subunit of *O. nova* TEBP. Furthermore, our biochemical studies demonstrated that, although itself does not bind ssDNA, TPP1 enhances the POT1-ssDNA interaction, exhibiting properties closely resembling those of TEBPβ [100]. Thus, our data clearly showed that TPP1 is the missing human homologue of the

O. nova TEBP β subunit and that capping of telomeres by a TEBPα/β dimer is more conserved evolutionary than had been expected. These results will be discussed in detail in Chapter 2.

1.6.2 Telomeric proteins that recognize the double-stranded telomeric DNA

Just as the OB-fold is the signature of all single-stranded G-overhang binding proteins, the Myb motif is found in all telomere binding proteins that recognize double-stranded telomere DNA. The Myb motif is named after the transcription factor, c-Myb, a proto-oncogene products that regulates differentiation and proliferation during hematopoiesis [101]. The Myb motif consists of three α-helices arranged in an orthogonal bundle around a hydrophobic core (Fig. 1.6). The third helix presents residues that make sequence-specific contacts with bases in the major groove of B-form DNA [102-104]. For telomere proteins, these DNA recognition residues are especially well conserved and defined the so-called telobox sequence feature [105, 106].

Budding yeast *Saccharomyces cerevisiae* Rap1 (repressor-activator protein 1) was the first identified double-stranded telomeric DNA-binding protein. Rap1 functions as a multifunctional protein that plays a crucial role in regulating telomere length, activating or repressing transcription [107-109]. Rap1 binds to the budding yeast irregular telomeric sequence (GTG_{1-3}) directly via its two tandem Myb motifs in the center of the protein [110, 111]. Rap1 is a negative regulator of telomere length [107, 108], and the number of bound Rap1 molecules at a telomere seems to constitute a telomere-length measuring mechanism [112, 113].

The double-stranded telomeric DNA in mammals is bound sequence specifically by two closely related proteins, TRF1 and TRF2. TRF1 (telomere repeat binding factor 1) was first identified from Hela cell extracts by using tandem TTAGGG repeats as the probe [114]. TRF1 contains three domains: an N-terminal acidic region, a dimerization domain in the center, and a DNA-binding Myb domain at C-terminus [115, 116]. TRF1 forms a dimer and remains a dimer when bound to the duplex TTAGGG repeats [116, 117]. The dimerization is mediated via the central domain that is referred to as TRF homology (TRFH) domain [118]. TRF1 functions as a negative regulator of telomere length in telomerase-positive cell lines. Over-expression of TRF1 leads to telomere length shortening whereas removal of TRF1 from telomeres causes progressive elongation of the telomere length [119]. Not long after the discovery of TRF1, a novel protein TRF2, which shares high homology with TRF1, especially the C-terminal Myb-domain sequence, was found in human and mouse [120, 121]. Structurally, both TRF1 and TRF2 contain a Myb-domain at the C-terminus that is responsible for interacting with the TTAGGG telomeric dsDNA directly. TRF2 also has the central TRFH domain that mediates the homodimer formation. The N-terminal of TRF2 is a basic region that prevents telomeres from being engaged in homologous recombination (HR) mediated telomere attrition [122]. TRF2 functions as a recruiter to bring other protein factors to telomeres [123]. Although TRF2 is also involved in telomere length regulation [124], its primary role appears to be in capping and protecting chromosome ends (reviewed in [125]).

The crystal structures of the TRFH domains from both human TRF1 and TRF2 revealed that they have almost identical and entirely α-helical dimeric structures [118].

The overall structure of the TRHF domain resembles a twisted horseshoe and each monomer contains 10 α helices. The dimer interface is formed by six α helices, α1, α2 and α10 form each monomers (Fig. 1.7). The biological importance of a dimeric structure of the TRF proteins was highlighted by point mutations at the dimeric interface that prevent telomere localization *in vivo* [118]. Despite the structural conservation, the sequence divergence between TRF1 and TRF2 TRFH domains results in different dimer interfaces that prevent TRF1/TRF2 heterodimer formation. Consideration of domain architecture in Rap1, TRF1 and TRF2 supports the view that telomere double-stranded DNA-binding proteins are constrained to use two Myb motifs for recognition of telomeric DNA. The two Myb units can be arranged either as tandem repeats as seen in *Sc*Rap1 or as individual units as seen in TRF1 and TRF2 that are brought together by protein dimerization.

Taz1 (telomere-associated protein in *Schizosaccharomyces pombe*) was identified as the double-stranded telomeric DNA-binding protein by using the one-hybrid screening in fission yeast [118, 126, 127]. Taz1 is involved in fission yeast telomere end protection, telomere length regulation, and 3' overhang processing [128, 129]. Deletion of $taz1^{+}$ in a $trt1^{+}$ background (TRT1 is the catalytic protein subunit of *S. pombe* telomerase), results in dramatically increased telomere length as well as length hererogeneity [126, 130]. It has been proposed that by involving in a negative feedback control of synthesis of its own binding sites, Taz1 could count the telomeric repeats, and thus function in the similar "protein-counting" mechanism as budding yeast Rap1 and human TRF1 [113, 119, 131]. In addition, similar to the TRF2-RAP1 interaction at human telomeres, Taz1 also recruits Rap1 (the fission yeast ortholog of human RAP1) to telomeres by direct

interaction [132, 133]. Structurally, Taz1 forms a homodimer and binds to the fission yeast telomeric double-stranded DNA in a sequence-specific manner [106, 134]. Taz1 contains a Myb domain at the C-terminus, which shares high sequence identity (29%) and similarity (66%) with TRF1 and TRF2 Myb domains [118, 126, 134]. Notably, Taz1 has a central predicted α-helical region (residues 113-400) that has low sequence similarity and identity to the TRFH domains of human TRF1 and TRF2 [118]. Nevertheless, it was proposed that this region contains the TRFH domain and mediates homodimer formation in a similar way as TRF1/TRF2-TRFH [127]. Collectively, functional and biochemical studies all suggested that Taz1 is the fission yeast ortholog of mammalian TRF1 and TRF2. Similar TRF-like dsDNA binding protein is also found in *Trypanosoma brucei* [135], which suggests that TRFH domain might be an ancient component of the telomeric complex and could be another conserved feature in most dsDNA binding telomeric proteins.

However, several key issues regarding Taz1 remains to be answered. Is the putative TRFH domain of Taz1 structurally similar to TRF1/TRF2-TRFH? Is this region indeed the dimerization domain of Taz1? How does Taz1 bind to Rap1 and does this interaction resemble the one between human TRF2 and RAP1? In Chapter 3, we will use both structural and functional approaches to address these questions.

1.6.3 The telomere shelterin complex

The core mammalian telomeric complex is called shelterin, for its role in telomere protection [136-138] (reviewed in [139]). It is delivered to telomeres by two of its components, TRF1 and TRF2. TRF2 recruits an additional factor, RAP1 [127]. The

mammalian G-overhang binding unit is a heterodimer of POT1-TPP1 [63]. A sixth factor, TRF1-interacting nuclear factor 2 (TIN2), interacts with three shelterin proteins, TRF1, TRF2 and TPP1, and thus has an important architectural role in bridging the duplex binding proteins, TRF1 and TRF2-RAP1, with the POT1-TPP1 complex at the single-stranded G-overhang (Fig. 1.8). Sheltein is at telomeres throughout the cell cycle, it does not accumulate elsewhere in the nucleus, and its function is limited to telomeres. This is the major difference that distinguishes shelterin from several DNA damage-processing factors and various other proteins that are also found at telomeres. These telomere-associated proteins can have crucial roles at telomeres, but also have nontelomeric functions. They often accumulate elsewhere in the cell, they are less abundant at telomeres than shelterin, and several are at telomeres transiently.

Shelterin plays an important role in keeping telomeres away from DNA damage checkpoints and therefore protects chromosome ends from inappropriate DNA repair pathways [139]. In addition, shelterin affects the structure of the telomere terminus and controls the synthesis of telomeric DNA by telomerase. POT1 is implicated in telomerase recruitment and C-strand processing [67, 86, 87]. The different possible localizations of POT1 at the 3’ G-overhang may switch telomerase on and off by occupying different positions on the substrate DNA for telomerase extension [62]. At the duplex region, TRF1, TRF2 and TIN2 are part of the negative feedback loop that regulates telomere length. RAP1 may function as a negative regulator for telomere length (reviewed in [32]). But, recent data suggested that RAP1 also plays an important role in telomere end protection [140]. Collectively, shelterin helps maintain the proper length and structure of telomeres and protect chromosome ends from DNA damage surveillance and repair

pathways.

A multi-protein telomeric complex with a shelterin-like architectural organization has been revealed in fission yeast *S. pombe* [141]. There are seven components in this complex, and many of them are the structural and functional homologues of the shelterin proteins (Fig. 1.9). Taz1 binds to the double-stranded telomeric DNA, regulates telomere homeostasis, and recruits another component Rap1 to telomeres. Thus, Taz1 functions as the *S. pombe* homolog of mammalian TRF1 and TRF2. At single-stranded G-overhangs, Pot1 and a Pot1-interacting partner, Tpz1, form a heterodimer that is the homolog of POT1-TPP1. Consistent with this notion, Pot1-Tpz1 protects telomeres and regulates telomerase activity [141]. Poz1, a small protein with no obvious sequence similarity to any components of the mammalian shelterin, interacts with both Tpz1 and Rap1, and thus connects the single-stranded and double-stranded binding proteins together (Fig. 1.9). This bridging function of Poz1 closely resembles the architectural role of TIN2 in the shelterin complex, raising the possibility that Poz1 might be a TIN2 homolog. One important difference is that Poz1 in *S. pombe*, unlike TIN2 in mammalian, interacts with Rap1. Another novel component is coiled-coil quantitatively enriched protein 1 (Ccq1). Ccq1 interacts with Tpz1 and plays a key role in recruiting telomerase to telomeres ([141], reviewed in [142]). Further work is required to identify the telomeric complexes in other organisms that, with no doubt, will bring us more mechanistic insights in the functions of these essential complexes.

1.7 Telomere higher-order structure, the t-loop

Evidence from a large number of studies suggests that both yeast and mammalian

telomeres exist in higher-order structures. A lasso-like structure called the t-loop was isolated from human telomeres from DNA cross-linking carried out *in vivo* and subsequently visualized by electron microscopy [143] (Fig. 1.10). The t-loop contains thousands of base pairs of TTAGGG repeats and is proposed to form by the insertion of the G-overhang into the double-stranded region of telomeric DNA [143]. During the invasion, a D-loop structure is also formed (Fig. 1.10). The t-loop can be reconstituted by incubating purified human TRF2 with double-stranded telomeric DNA ending in a single-stranded G-overhang [144]. Subsequently, t-loop structures have been observed at the end of telomeres of many other organisms including mouse, chicken, pea, trypanosome and yeast mitochondria, and hence appear to be an evolutionarily conserved feature of telomere structure.

Repetitive telomeric DNA and a homologous single-stranded 3' G-overhang are structurally essential for t-loops to form. However, these elements also make the telomere a prime target for homologous recombination. The t-loop is structural mechanism that sequesters the ends of linear chromosomes, prevents recognition of the chromosome end as a double strand DNA break and inhibits exonulceolytic degradation; but the t-loop itself poses another problem. In particular, t-loops resemble homologous recombination intermediates and branch migration of a t-loop could generate a Holliday junction. This raises the question, how do telomeric DNA and its associated proteins prevent inappropriate recombination at the telomere or t-loop? Recent data suggest that double-stranded telomeric DNA binding protein TRF2 and a newly identified BLM complex play important roles in telomere protection and maintenance. The structural study of the BLM complex is discussed in Chapter 4.

1.8 Telomere, telomerase and cancer

Telomerase is highly expressed in cells that need to divide regularly. In embryonic stem cells and specific germ-line cells, telomerase is expressed and remains fully active, [48, 49, 145, 146] (reviewed in [32, 147]). But, it is undetectable in most normal somatic cells except for proliferative cells of renewal tissues. Therefore, telomere length is shortened progressively in these nonproliferative somatic cells. Telomere shortening is hypothesized to be the direct cause for the "Hayflick limit" and a mechanism that prevents normal cells from developing to cancerous cells [148]. In 1998, Bodnar *et al.* demonstrated the relationship between telomere shortening and proliferative failure in human cells [149]. Transfection of hTERT into two telomerase-negative normal human cell lines successfully restored the telomerase activity. These hTERT transfected cells exhibited longer telomeres, indefinite life span and lower level of senescence.

While telomerase activity is undetectable in most normal somatic cells, high levels of telomerase expression or re-expression and robust telomerase activity can be detected in more than 90% of tumor samples [150-156]. This discovery suggested that telomerase is a very good marker for cancer diagnosis and a potential target for anti-cancer therapy [157-159]. The regulation of telomerase activity usually occurs at the gene transcription level. The transcription of *hTERT* gene is repressed by numerous factors, thus affecting telomerase assembly [160, 161] (reviewed in [161]). A genetic screen searching for the negative regulators of *hTERT* was performed by Lin and Elledge and identified three tumor suppressor/oncogene pathways that have the ability to repress *hTERT*, *Mad1/c-Myc* pathway (implicate in *hTERT* regulation), *SIP1* (a transcriptional

target of *TGF-* pathway, mediates *TGF-*–regulated *hTERT* suppression) and Menin (a direct *hTERT* repressor) [162]. Although tumor cells require telomerase activity for keeping long-term proliferation, the *hTERT* gene itself is not an oncogene and is not sufficient to induce cell transformation [163, 164]. Thus, telomerase is controlled at a subtle balance between cellular senescence and immortalization.

Not all the cancer cells and tumor tissues require telomerase to maintain the potential for proliferation. A telomerase-independent telomere maintenance pathway has been identified and named the “alternative lengthening of telomeres” (ALT) mechanism [165]. The available data indicate that ALT involves homologous recombination-mediated DNA replication and requires the activity of many recombination complexes. Increased levels of various types of telomere recombination events in ALT cells suggest that the cellular mechanisms that normally regulate recombination at mammalian telomeres have been lost. Understanding how telomeres are maintained in ALT cells will help reveal the mechanisms that have evolved in mammalian cells (primary cells) to inhibit deregulated homologous recombination at the telomeres and thus prevent telomere elongation and cellular immortalization.

Figure 1.1 End replication problem
The leading strand (green) is continuously synthesized from 5’ to 3’ by polymerase. The lagging strand (blue) synthesis is initialized by RNA primers (red) from 3’ to 5’. When RNA primers are removed, a gap will be left at the 3’ end.

End Replication Problem

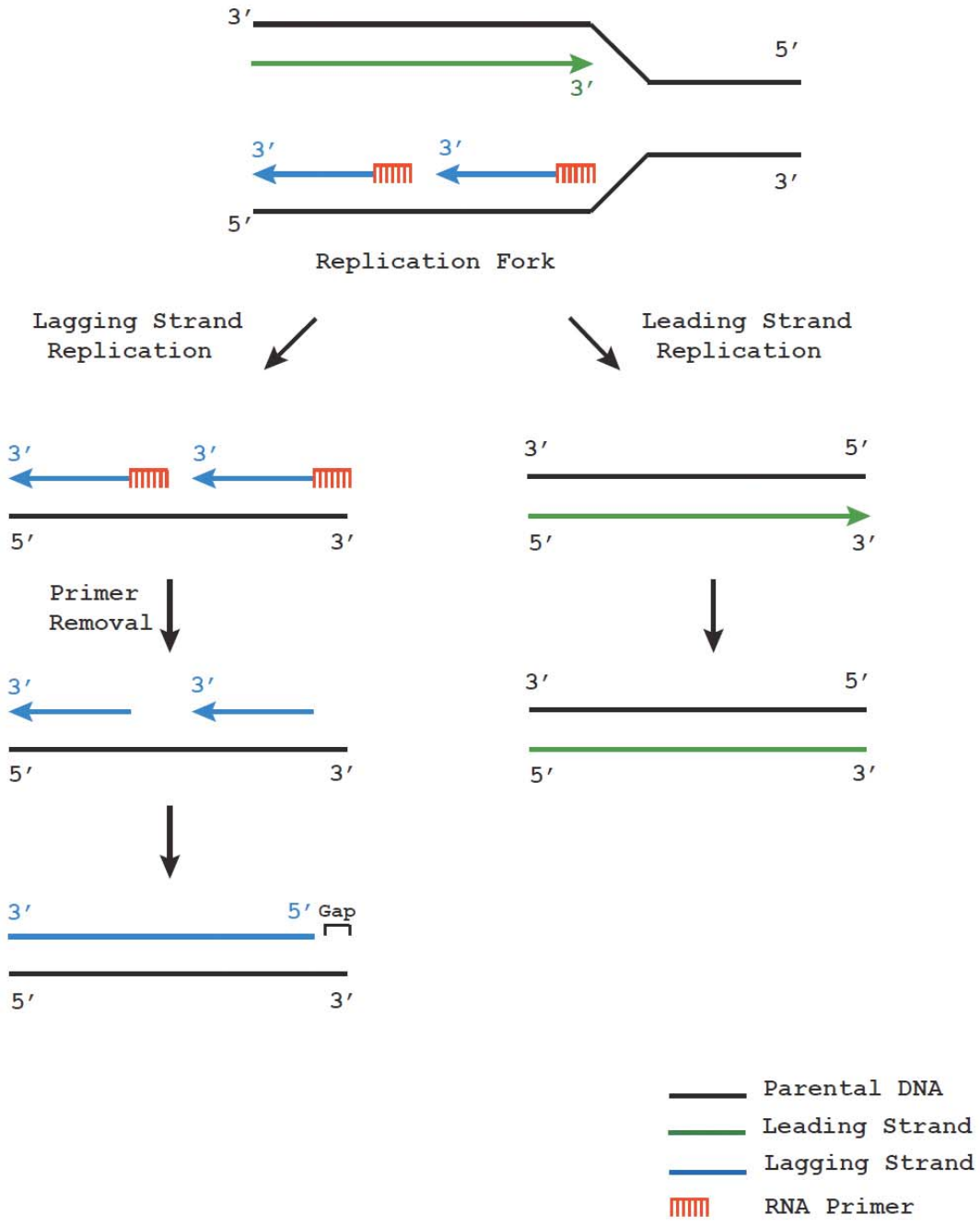


Figure 1.2* Cavalier-Smith's model for the synthesis of 5' ends of a daughter DNA molecule (the 3' end of a polynucleotide shown by an arrow, cyan and green lines are two complementary strands, newly synthesized DNA by a thick line).
(a) Gap at 5' end of new DNA;
(b) The self complementary 3' terminal strand of the palindromic sequence base pairs to form a hairpin loop;
(c) 3' OH end serves as a primer for a DNA polymerase to fill in the gap.
(d) DNA ligase seals the remaining nick;
(e) A sequence specific endonuclease nicks the old strand;
(f) The loop unfolds;
(g) DNA polymerase completes the old strand with new DNA.
(* This figure is made based on ref. [20]Fig. 2)

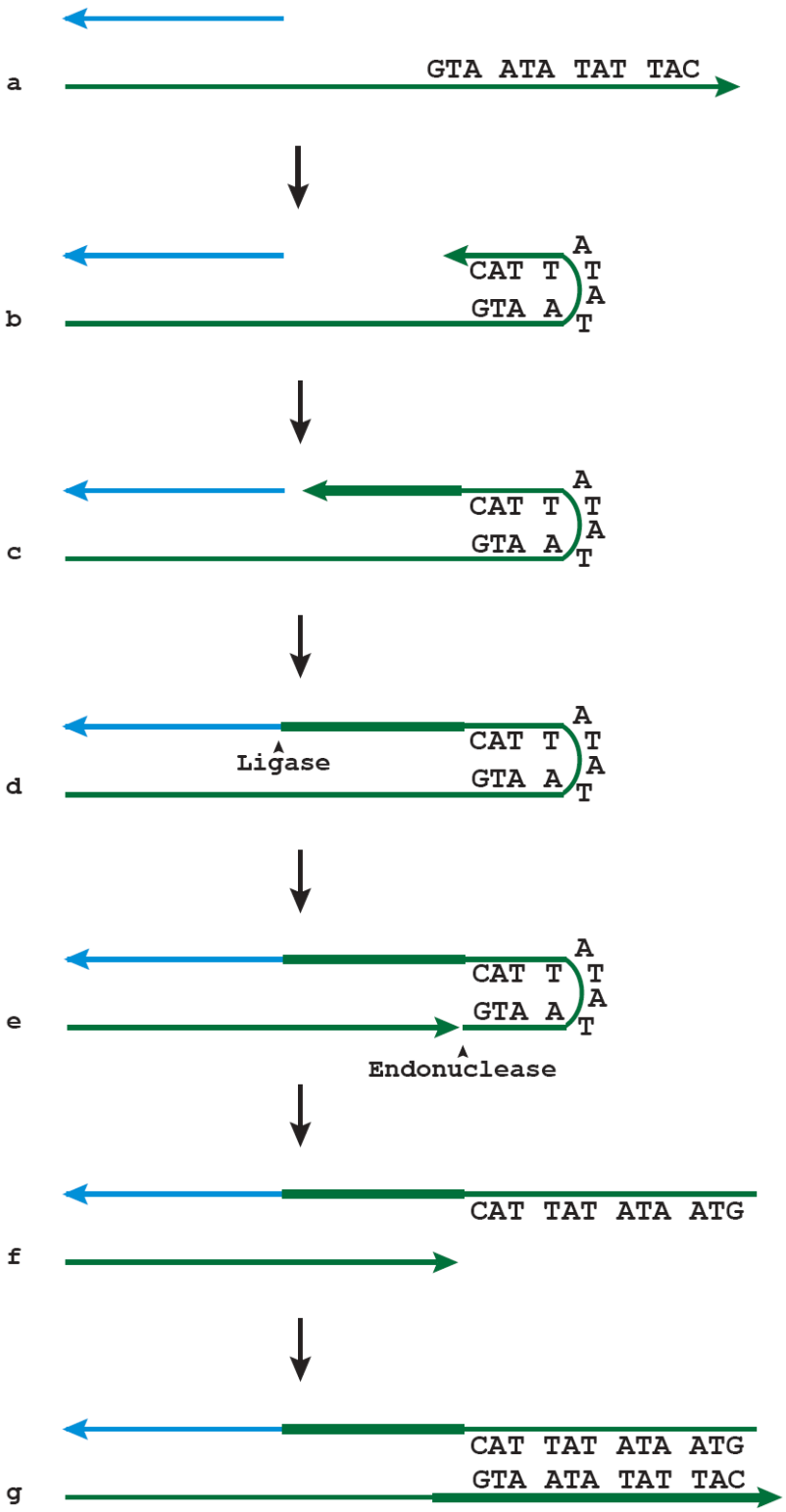

a
GTA ATA TAT TAC
b
A
CAT T T
A
GTA A T
c
A
CAT T T
A
GTA A T
d
Ligase
A
CAT T T
A
GTA A T
e
A
CAT T T
A
GTA A T
Endonuclease
f
CAT TAT ATA ATG
g
CAT TAT ATA ATG
GTA ATA TAT TAC

TABLE 1.1 Telomeric DNA Sequence

Species	Organism	5' - Telomeric DNA Repeat Sequence - 3'
Vertebrates	Human, mouse, *Xenopus*	TTAGGG
Higher plants	*Arabidopsis thaliana*	TTTAGGG
Budding yeast	*Saccharomyces cerevisiae*	$G_{(2-3)}(TG)_{(1-6)}T$
	Candida albicans	GGTGTACGGATGTCTAACTTCTT
	Candida glabrata	GGGGTCTGGGTGCTG
	Candida guillermondii	GGTGTAC
	Candida tropicalis	GGTGTA[C/A]GGATGTCACGATCATT
Fission yeast	*Schizosaccharomyces pombe*	$TTAC(A)(C)G_{(1-8)}$
Ciliate protozoa	*Tetrahymena*	TTGGGG
	Oxytricha	TTTTGGGG
Kinetoplastid protozoa	*Trypanosoma*	TTAGGG

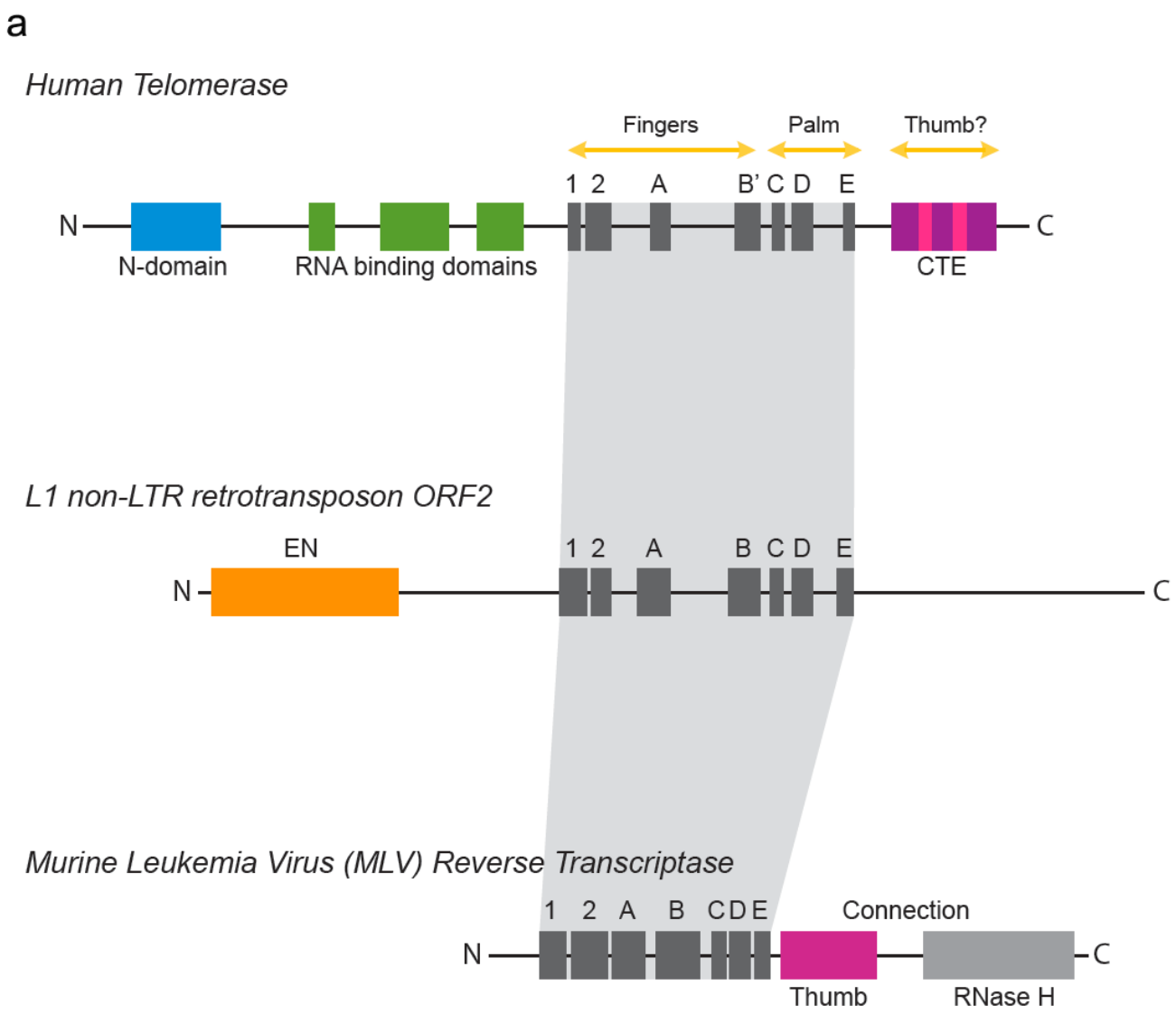
Figure 1.3* Structure illustrations of human TERT and TR

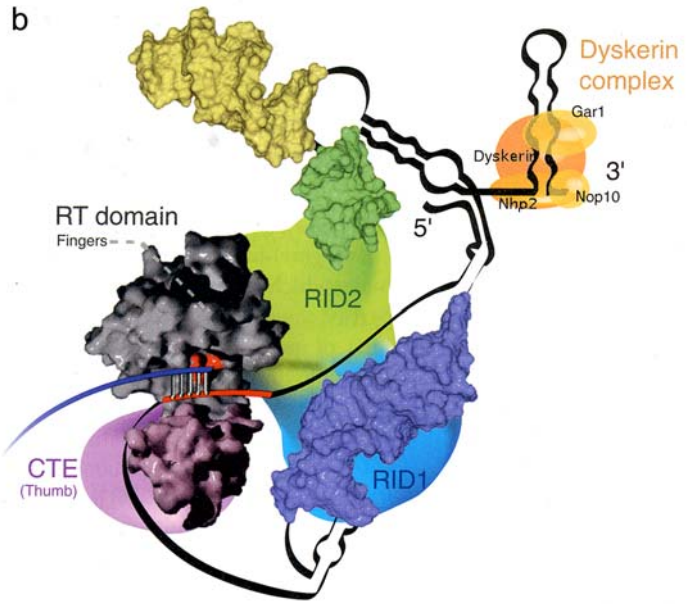
(a) Domain organization of human TERT and comparison to RTs from retroviruses and retrotransposons. The RT domain (dark grey) includes conserved RT motifs 1 to E (fingers and palm) and the thumb domain, which is much more divergent among RTs (pink). The CTE (carboxyl-terminal extension) of telomerases contains three blocks of conserved amino acids. The amino-terminal half contains one or two RNA-binding domains (green). The region in blue may bind to another region in TR. Endonuclease (EN, orange) and RNase H (light gray) domains found in some retrotransposons and retroviruses have not been identified in TERT.

(b) Model of a monomeric human telomerase RNP in the absence of hEST1A/B and other telomerase-associated proteins. The central RT structure (grey and pink) corresponds to murine leukemia virus RT [52] and includes its finger, palm, and thumb domains but not the RNase H domain. The colors are the same as those in **(a)**.

(c) Telomerase RNA secondary structure model for *Homo sapiens*. Different regions are highlighted by different color, as indicated.

(* Figure 1.3a is made based on Fig. 2 from Chapter 2 The Telomerase Ribonucleoprotein Particle, *Telomeres* 2nd edition 2006, Cold Spring Harbor Laboratory Press. Figure 1.3b and 1.3c are the photocopies of Fig. 3 and Fig. 1B from Chapter 2 The Telomerase Ribonucleoprotein Particle, *Telomeres* 2nd edition 2006, Cold Spring Harbor Laboratory Press.)


a
Human Telomerase
Fingers
Palm
Thumb?
1 2 A B' C D E
N
C
N-domain
RNA binding domains
CTE
L1 non-LTR retrotransposon ORF2
EN
1 2 A B C D E
N
C
Murine Leukemia Virus (MLV) Reverse Transcriptase
1 2 A B C D E
Connection
N
C
Thumb
RNase H


b
Dyskerin complex
Gar1
Dyskerin
3'
Nop10
Nhp2
5'
RT domain
Fingers
RID2
CTE
(Thumb)
RID1

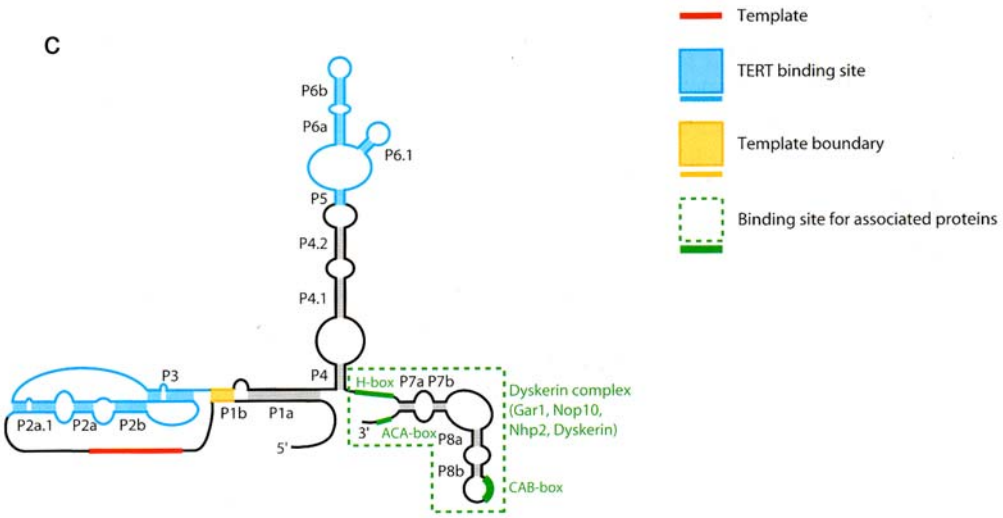
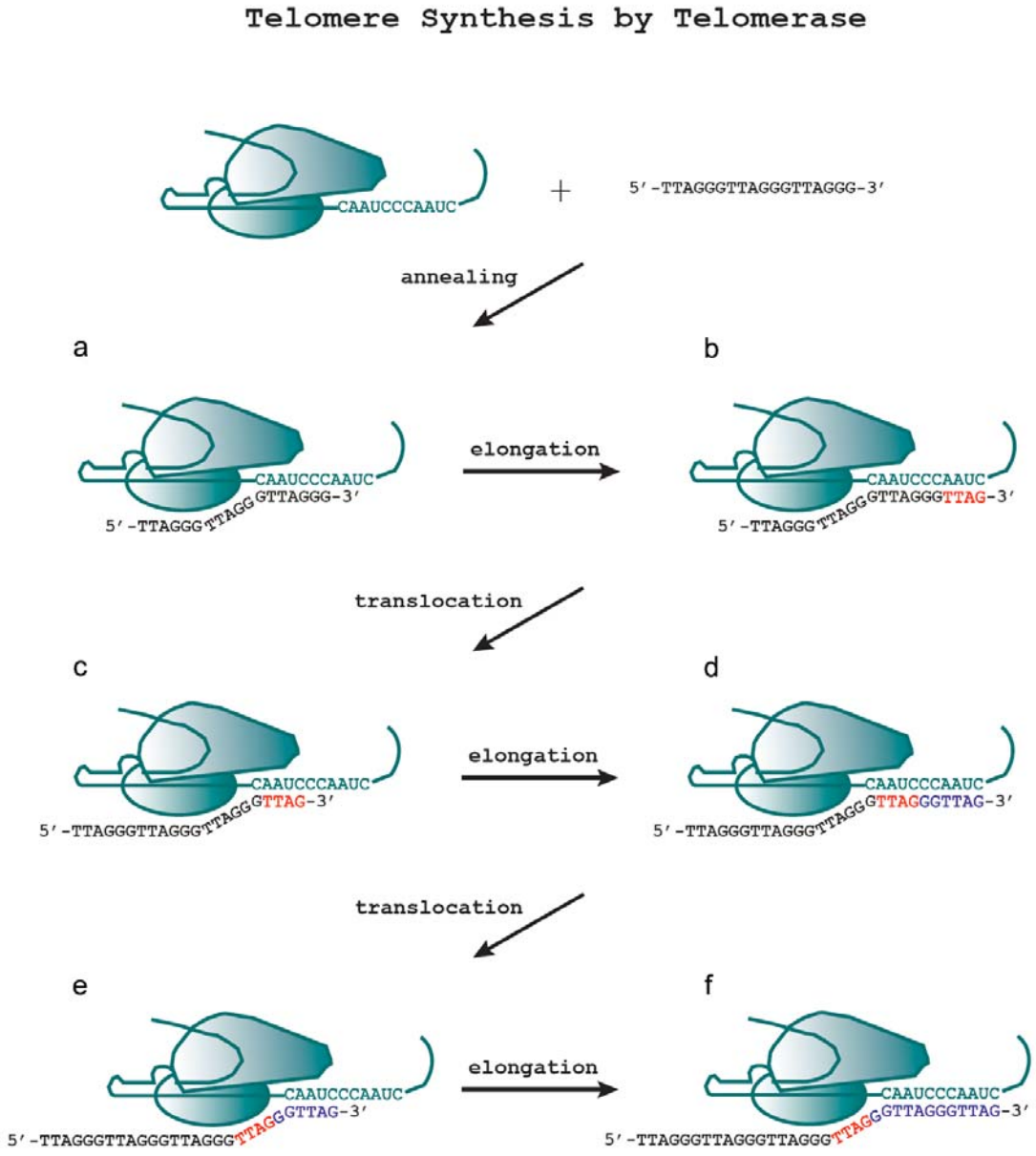

c
Template
TERT binding site
Template boundary
Binding site for associated proteins
P6b
P6a
P6.1
P5
P4.2
P4.1
P3
P4
H-box
P7a P7b
Dyskerin complex
(Gar1, Nop10,
Nhp2, Dyskerin)
P2a.1
P2a
P2b
P1b
P1a
5'
3'
ACA-box
P8a
P8b
CAB-box

Figure 1.4 Telomerase elongates telomere in a processive manner.
(a) Telomerase anneals to telomere via its RNA template;
(b) Telomerase adds nucleotides (red) to the end of telomere;
(c) Telomerase translocates to the next binding site without falling off the telomere;
(d) Repeats step **(b)**, the added nucleotides are colored in blue;
(e) Repeats step **(c)**
(f) Repeats step **(b)**, the added nucleotides are colored in blue.


Telomere Synthesis by Telomerase
CAAUCCCAAUC
5′-TTAGGGTTAGGGTTAGGG-3′
annealing
a
b
elongation
CAAUCCCAAUC
5′-TTAGGGTTAGGGTTAGGG-3′
CAAUCCCAAUC
5′-TTAGGGTTAGGGTTAGGGTTAG-3′
translocation
c
d
elongation
CAAUCCCAAUC
5′-TTAGGGTTAGGGTTAGGGTTAG-3′
CAAUCCCAAUC
5′-TTAGGGTTAGGGTTAGGGTTAGGGTTAG-3′
translocation
e
f
elongation
CAAUCCCAAUC
5′-TTAGGGTTAGGGTTAGGGTTAGGGTTAG-3′
CAAUCCCAAUC
5′-TTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAG-3′

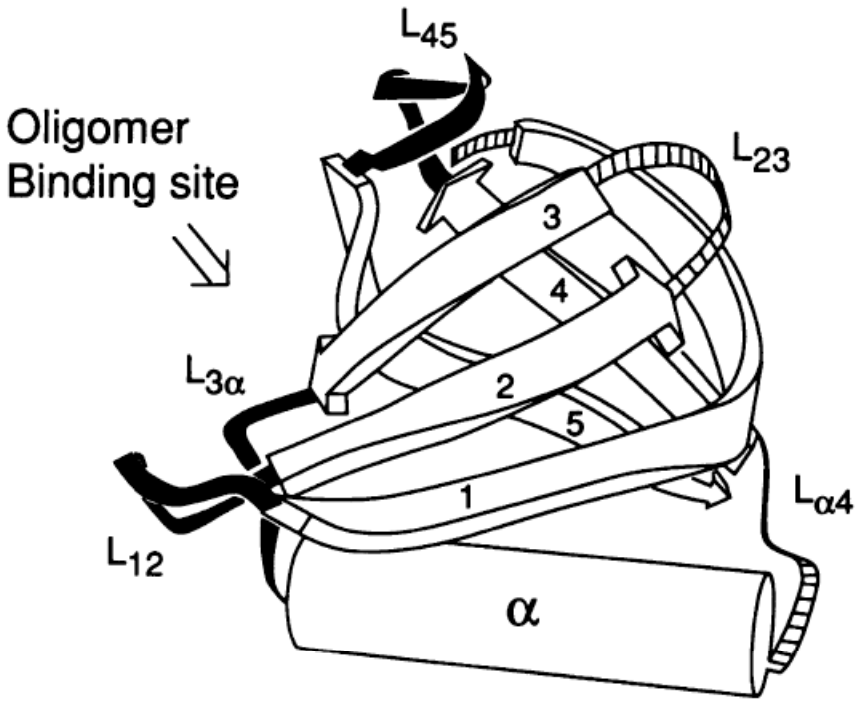


Figure 1.5 Cartoon illustration of OB-fold Structural architecture of a typical OB-fold (ref [94], Fig. 1); a typical OB-fold has the β-β-β-α-β-β secondary structural arrangement. The arrow indicates the oligomer binding site.

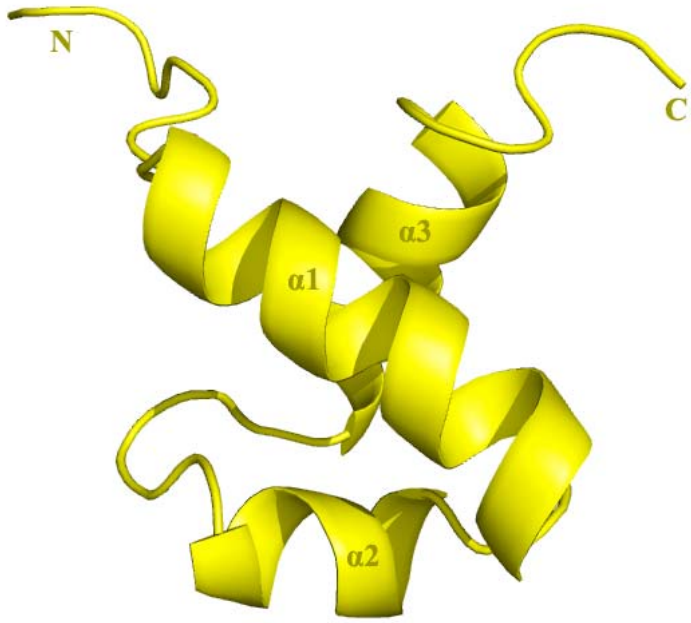


Figure 1.6 Structure of a typical c-Myb domain Cartoon illustration of the mouse c-Myb DNA-binding domain repeat 3 (PDB#: 1idy). Myb-domain usually contains three α helices.

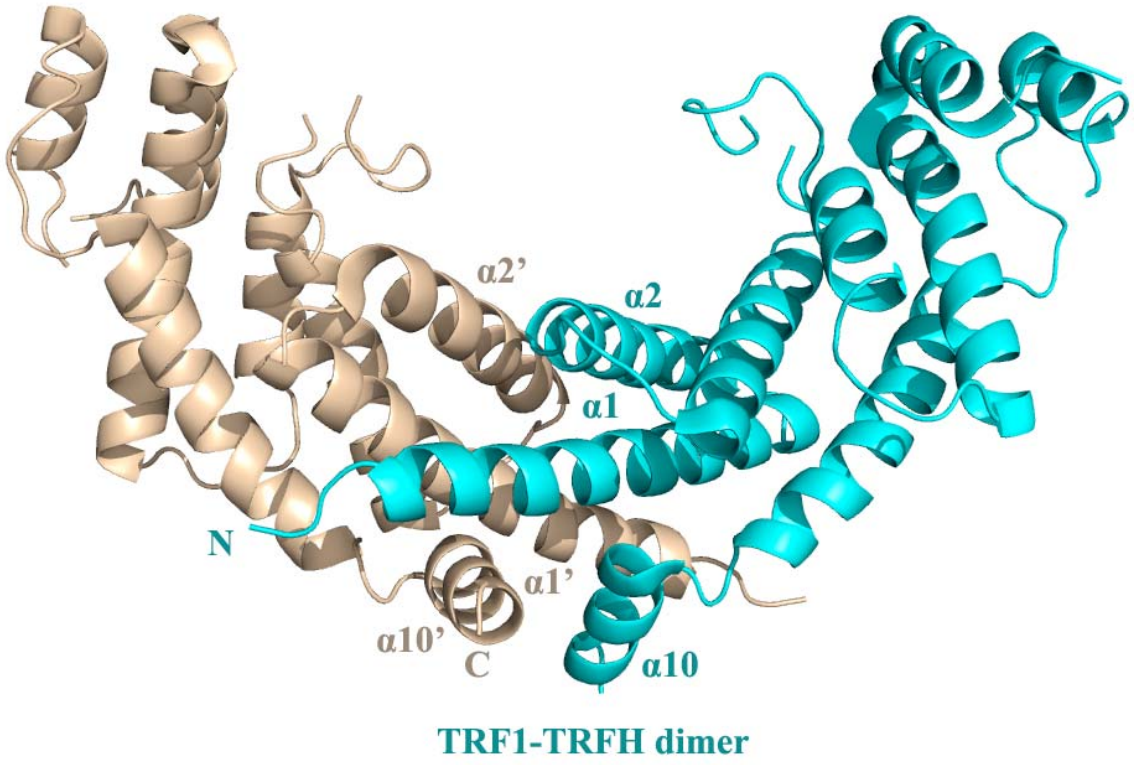
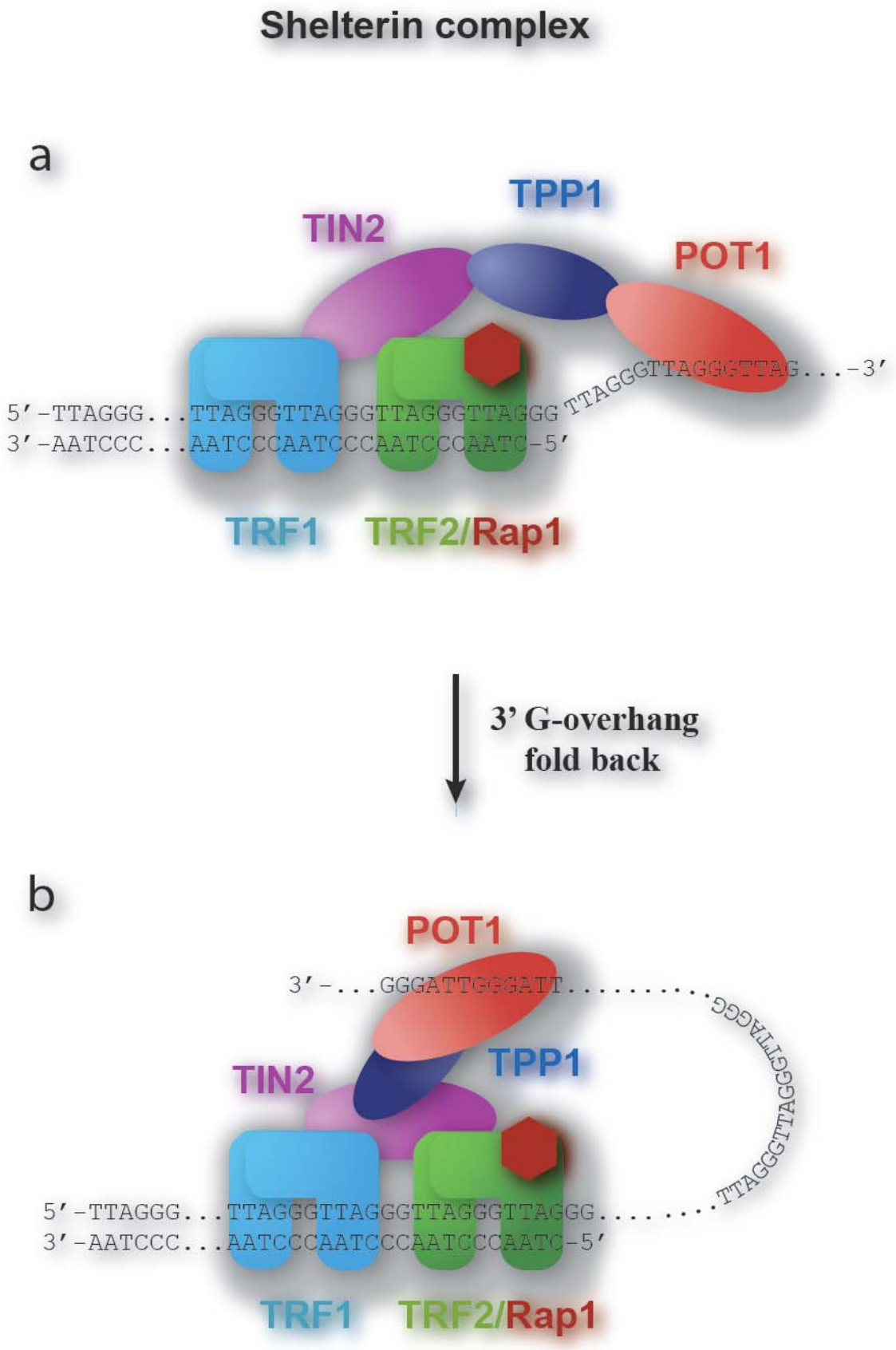


Figure 1.7 Crystal structure of human TRF1-TRFH dimer Cartoon illustration of the TRFH domain of human double-stranded telomeric DNA binding protein TRF1 (PDB#: 1H6O). TRF1-TRFH forms a homodimer through the interactions between six α helices, α1, α2 and α10 from each of the two monomers.

Figure 1.8 Cartoon illustration of shelterin complex
(a) Six shelterin complex components: TRF1 (cyan) and TRF2 (green) on the double-stranded telomeric DNA; TIN2 (purple) binds to both TRF1 and TRF2; Rap1 (light green) interacts with TRF2; TPP1 (blue) interacts with both TIN2 and POT1 (red). POT1 binds to single-stranded telomeric DNA.
(b) 3' G-overhang is folded back due to TPP1-TIN2 interaction.


Shelterin complex
a
TPP1
TIN2
POT1
5′-TTAGGG...TTAGGGTTAGGGTTAGGGTTAGGG TTAGGGTTAGGGTTAG...-3′
3′-AATCCC...AATCCCAATCCCAATCCCAATC-5′
TRF1
TRF2/Rap1
3’ G-overhang
fold back
b
POT1
3′-...GGGATTGGGATT.....GGGTTAGGG
TPP1
TIN2
5′-TTAGGG...TTAGGGTTAGGGTTAGGGTTAGGG.....TTAGGGTTAGGG
3′-AATCCC...AATCCCAATCCCAATCCCAATC-5′
TRF1
TRF2/Rap1

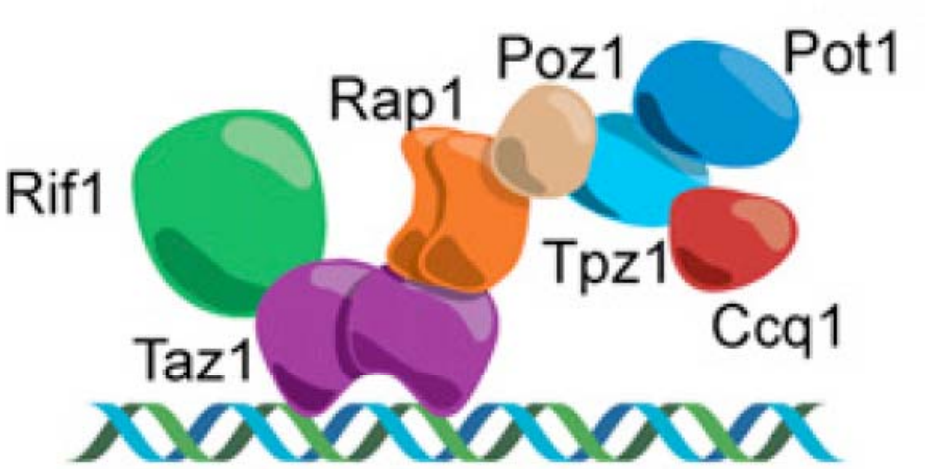
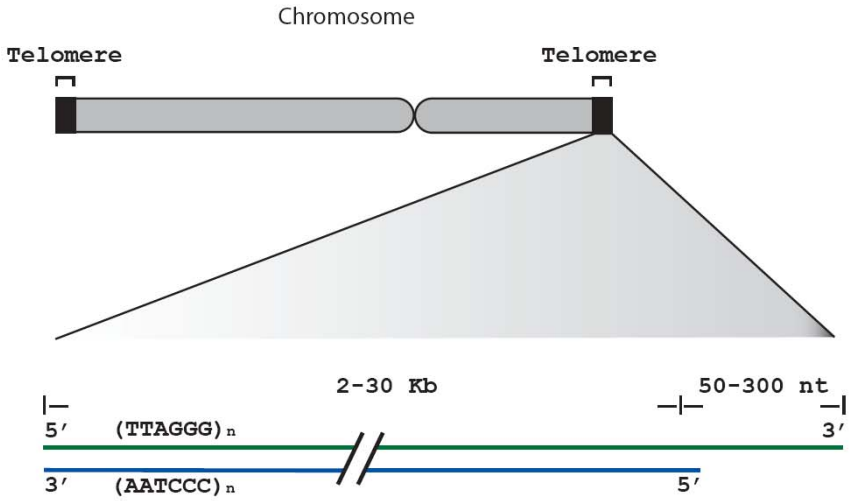


Figure 1.9* The shelterin-like complex in fission yeast *S. pombe* There are seven components in this complex: Taz1 (purple) on the duplex telomeric DNA, Rif1 (green) and Rap1 (orange) interact with Taz1. Pot1 (blue) binds to single-stranded telomeric DNA (not shown in this figure) and interacts with Tpz1 (cyan). Ccq1 (dark red) interacts with Tpz1. Poz1 (wheat) connects Tpz1 and Rap1. Some of them are the structural and functional homologues of the shelterin proteins.
(* This figure is photocopied from Fig. 2 of the review paper by Bianchi and Shore. [142])

Figure 1.10 Telomere and T-loop
(a) Telomere region is divided into 2 regions: double-stranded telomeric DNA region and 3’ G-overhang. Green and cyan lines are two complementary strands of DNA.
(b) 3’ G-overhang folds back and invades into the double-stranded DNA track. D-loop and T-loop are formed.

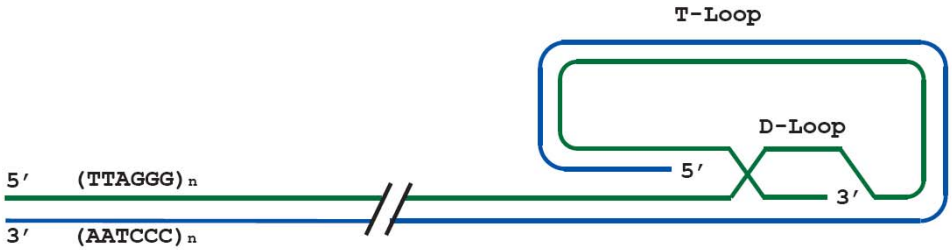
Telomere and T-loop

a



double-stranded telomeric DNA region **3' G-overhang**

b



Reference:

1. McClintock, B., *The Behavior in Successive Nuclear Divisions of a Chromosome Broken at Meiosis.* Proc Natl Acad Sci U S A, 1939. 25(8): p. 405-16.
2. Hayflick, L. and P.S. Moorhead, *The serial cultivation of human diploid cell strains.* Exp Cell Res, 1961. 25: p. 585-621.
3. Hayflick, L., *The Limited in Vitro Lifetime of Human Diploid Cell Strains.* Exp Cell Res, 1965. 37: p. 614-36.
4. Hayflick, L., *A brief history of the mortality and immortality of cultured cells.* Keio J Med, 1998. 47(3): p. 174-82.
5. Olovnikov, A.M., *A theory of marginotomy. The incomplete copying of template margin in enzymic synthesis of polynucleotides and biological significance of the phenomenon.* J Theor Biol, 1973. 41(1): p. 181-90.
6. Cooke, H.J. and B.A. Smith, *Variability at the telomeres of the human X/Y pseudoautosomal region.* Cold Spring Harb Symp Quant Biol, 1986. 51 Pt 1: p. 213-9.
7. Greider, C.W. and E.H. Blackburn, *Identification of a specific telomere terminal transferase activity in Tetrahymena extracts.* Cell, 1985. 43(2 Pt 1): p. 405-13.
8. Meselson, M. and F.W. Stahl, *The Replication of DNA in Escherichia Coli.* Proc Natl Acad Sci U S A, 1958. 44(7): p. 671-82.
9. Okazaki, R., et al., *Mechanism of DNA replication possible discontinuity of DNA chain growth.* Jpn J Med Sci Biol, 1967. 20(3): p. 255-60.
10. Ogawa, T. and T. Okazaki, *Discontinuous DNA replication.* Annu Rev Biochem, 1980. 49: p. 421-57.
11. Ohki, R., T. Tsurimoto, and F. Ishikawa, *In vitro reconstitution of the end replication problem.* Mol Cell Biol, 2001. 21(17): p. 5753-66.
12. Baird, D.M., *Telomere dynamics in human cells.* Biochimie, 2008. 90(1): p. 116-21.
13. Watson, J.D., *Origin of concatemeric T7 DNA.* Nat New Biol, 1972. 239(94): p. 197-201.
14. Broker, T.R., *An electron microscopic analysis of pathways for bacteriophage T4 DNA recombination.* J Mol Biol, 1973. 81(1): p. 1-16.
15. Challberg, M.D. and T.J. Kelly, Jr., *Adenovirus DNA replication in vitro: origin and direction of daughter strand synthesis.* J Mol Biol, 1979. 135(4): p. 999-1012.
16. Challberg, M.D., S.V. Desiderio, and T.J. Kelly, Jr., *Adenovirus DNA replication in vitro: characterization of a protein covalently linked to nascent DNA strands.* Proc Natl Acad Sci U S A, 1980. 77(9): p. 5105-9.
17. Desiderio, S.V. and T.J. Kelly, Jr., *Structure of the linkage between adenovirus DNA and the 55,000 molecular weight terminal protein.* J Mol Biol, 1981. 145(2): p. 319-37.
18. Alestrom, P., et al., *DNA sequence analysis of the region encoding the terminal protein and the hypothetical N-gene product of adenovirus type 2.* J Biol Chem, 1982. 257(22): p. 13492-8.
19. Smart, J.E. and B.W. Stillman, *Adenovirus terminal protein precursor. Partial amino acid sequence and the site of covalent linkage to virus DNA.* J Biol Chem, 1982. 257(22): p. 13499-506.

20. Cavalier-Smith, T., *Palindromic base sequences and replication of eukaryote chromosome ends.* Nature, 1974. 250(5466): p. 467-70.
21. Gall, J.G., *Free ribosomal RNA genes in the macronucleus of Tetrahymena.* Proc Natl Acad Sci U S A, 1974. 71(8): p. 3078-81.
22. Prescott, D.M. and K.G. Murti, *Chromosome structure in ciliated protozoans.* Cold Spring Harb Symp Quant Biol, 1974. 38: p. 609-18.
23. Karrer, K.M. and J.G. Gall, *The macronuclear ribosomal DNA of Tetrahymena pyriformis is a palindrome.* J Mol Biol, 1976. 104(2): p. 421-53.
24. Engberg, J., et al., *Free ribosomal DNA molecules from Tetrahymena pyriformis GL are giant palindromes.* J Mol Biol, 1976. 104(2): p. 455-70.
25. Blackburn, E.H. and J.G. Gall, *A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena.* J Mol Biol, 1978. 120(1): p. 33-53.
26. Szostak, J.W. and E.H. Blackburn, *Cloning yeast telomeres on linear plasmid vectors.* Cell, 1982. 29(1): p. 245-55.
27. Oka, Y., et al., *Inverted terminal repeat sequence in the macronuclear DNA of Stylonychia pustulata.* Gene, 1980. 10(4): p. 301-6.
28. Klobutcher, L.A., et al., *All gene-sized DNA molecules in four species of hypotrichs have the same terminal sequence and an unusual 3' terminus.* Proc Natl Acad Sci U S A, 1981. 78(5): p. 3015-9.
29. Pluta, A.F., B.P. Kaine, and B.B. Spear, *The terminal organization of macronuclear DNA in Oxytricha fallax.* Nucleic Acids Res, 1982. 10(24): p. 8145-54.
30. Harley, C.B., et al., *The telomere hypothesis of cellular aging.* Exp Gerontol, 1992. 27(4): p. 375-82.
31. Forsyth, N.R., et al., *Developmental differences in the immortalization of lung fibroblasts by telomerase.* Aging Cell, 2003. 2(5): p. 235-43.
32. Smogorzewska, A. and T. de Lange, *Regulation of telomerase by telomeric proteins.* Annu Rev Biochem, 2004. 73: p. 177-208.
33. Gasser, S.M., *A sense of the end.* Science, 2000. 288(5470): p. 1377-9.
34. Blackburn, E.H., *Switching and signaling at the telomere.* Cell, 2001. 106(6): p. 661-73.
35. de Lange, T., *Protection of mammalian telomeres.* Oncogene, 2002. 21(4): p. 532-40.
36. Shampay, J., J.W. Szostak, and E.H. Blackburn, *DNA sequences of telomeres maintained in yeast.* Nature, 1984. 310(5973): p. 154-7.
37. Greider, C.W. and E.H. Blackburn, *The telomere terminal transferase of Tetrahymena is a ribonucleoprotein enzyme with two kinds of primer specificity.* Cell, 1987. 51(6): p. 887-98.
38. Greider, C.W. and E.H. Blackburn, *A telomeric sequence in the RNA of Tetrahymena telomerase required for telomere repeat synthesis.* Nature, 1989. 337(6205): p. 331-7.
39. Yu, G.L., et al., *In vivo alteration of telomere sequences and senescence caused by mutated Tetrahymena telomerase RNAs.* Nature, 1990. 344(6262): p. 126-32.

40. Lendvay, T.S., et al., *Senescence mutants of Saccharomyces cerevisiae with a defect in telomere replication identify three additional EST genes.* Genetics, 1996. 144(4): p. 1399-412.
41. Lingner, J. and T.R. Cech, *Purification of telomerase from Euplotes aediculatus: requirement of a primer 3' overhang.* Proc Natl Acad Sci U S A, 1996. 93(20): p. 10712-7.
42. Feng, J., et al., *The RNA component of human telomerase.* Science, 1995. 269(5228): p. 1236-41.
43. Singer, M.S. and D.E. Gottschling, *TLC1: template RNA component of Saccharomyces cerevisiae telomerase.* Science, 1994. 266(5184): p. 404-9.
44. Shippen-Lentz, D. and E.H. Blackburn, *Functional evidence for an RNA template in telomerase.* Science, 1990. 247(4942): p. 546-52.
45. Cohen, S.B., et al., *Protein composition of catalytically active human telomerase from immortal cells.* Science, 2007. 315(5820): p. 1850-3.
46. Harrington, L., et al., *A mammalian telomerase-associated protein.* Science, 1997. 275(5302): p. 973-7.
47. Lingner, J., et al., *Three Ever Shorter Telomere (EST) genes are dispensable for in vitro yeast telomerase activity.* Proc Natl Acad Sci U S A, 1997. 94(21): p. 11190-5.
48. Meyerson, M., et al., *hEST2, the putative human telomerase catalytic subunit gene, is up-regulated in tumor cells and during immortalization.* Cell, 1997. 90(4): p. 785-95.
49. Nakamura, T.M., et al., *Telomerase catalytic subunit homologs from fission yeast and human.* Science, 1997. 277(5328): p. 955-9.
50. Aubert, G. and P.M. Lansdorp, *Telomeres and aging.* Physiol Rev, 2008. 88(2): p. 557-79.
51. O'Reilly, M., S.A. Teichmann, and D. Rhodes, *Telomerases.* Curr Opin Struct Biol, 1999. 9(1): p. 56-65.
52. Das, D. and M.M. Georgiadis, *The crystal structure of the monomeric reverse transcriptase from Moloney murine leukemia virus.* Structure, 2004. 12(5): p. 819-29.
53. Leeper, T., N. Leulliot, and G. Varani, *The solution structure of an essential stem-loop of human telomerase RNA.* Nucleic Acids Res, 2003. 31(10): p. 2614-21.
54. Leeper, T.C. and G. Varani, *The structure of an enzyme-activating fragment of human telomerase RNA.* RNA, 2005. 11(4): p. 394-403.
55. Theimer, C.A., C.A. Blois, and J. Feigon, *Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function.* Mol Cell, 2005. 17(5): p. 671-82.
56. Gillis, A.J., A.P. Schuller, and E. Skordalakes, *Structure of the Tribolium castaneum telomerase catalytic subunit TERT.* Nature, 2008. 455(7213): p. 633-7.
57. Romero, D.P. and E.H. Blackburn, *A conserved secondary structure for telomerase RNA.* Cell, 1991. 67(2): p. 343-53.
58. Lingner, J., L.L. Hendrick, and T.R. Cech, *Telomerase RNAs of different ciliates have a common secondary structure and a permuted template.* Genes Dev, 1994. 8(16): p. 1984-98.

59. Chen, J.L., M.A. Blasco, and C.W. Greider, *Secondary structure of vertebrate telomerase RNA.* Cell, 2000. 100(5): p. 503-14.
60. Dandjinou, A.T., et al., *A phylogenetically based secondary structure for the yeast telomerase RNA.* Curr Biol, 2004. 14(13): p. 1148-58.
61. Zappulla, D.C. and T.R. Cech, *Yeast telomerase RNA: a flexible scaffold for protein subunits.* Proc Natl Acad Sci U S A, 2004. 101(27): p. 10024-9.
62. Lei, M., et al., *Switching human telomerase on and off with hPOT1 protein in vitro.* J Biol Chem, 2005. 280(21): p. 20449-56.
63. Wang, F., et al., *The POT1-TPP1 telomere complex is a telomerase processivity factor.* Nature, 2007. 445(7127): p. 506-10.
64. Greider, C.W., *Telomerase is processive.* Mol Cell Biol, 1991. 11(9): p. 4572-80.
65. Sun, D., et al., *Regulation of catalytic activity and processivity of human telomerase.* Biochemistry, 1999. 38(13): p. 4037-44.
66. Sfeir, A.J., et al., *Telomere-end processing the terminal nucleotides of human chromosomes.* Mol Cell, 2005. 18(1): p. 131-8.
67. Hockemeyer, D., et al., *POT1 protects telomeres from a transient DNA damage response and determines how human chromosomes end.* EMBO J, 2005. 24(14): p. 2667-78.
68. Henderson, E.R. and E.H. Blackburn, *An overhanging 3' terminus is a conserved feature of telomeres.* Mol Cell Biol, 1989. 9(1): p. 345-8.
69. McElligott, R. and R.J. Wellinger, *The terminal DNA structure of mammalian chromosomes.* EMBO J, 1997. 16(12): p. 3705-14.
70. Wright, W.E., et al., *Normal human chromosomes have long G-rich telomeric overhangs at one end.* Genes Dev, 1997. 11(21): p. 2801-9.
71. Gottschling, D.E. and V.A. Zakian, *Telomere proteins: specific recognition and protection of the natural termini of Oxytricha macronuclear DNA.* Cell, 1986. 47(2): p. 195-205.
72. Price, C.M. and T.R. Cech, *Telomeric DNA-protein interactions of Oxytricha macronuclear DNA.* Genes Dev, 1987. 1(8): p. 783-93.
73. Sheng, H., et al., *Identification and characterization of a putative telomere end-binding protein from Tetrahymena thermophila.* Mol Cell Biol, 1995. 15(3): p. 1144-53.
74. Lin, J.J. and V.A. Zakian, *The Saccharomyces CDC13 protein is a single-strand TG1-3 telomeric DNA-binding protein in vitro that affects telomere behavior in vivo.* Proc Natl Acad Sci U S A, 1996. 93(24): p. 13760-5.
75. Nugent, C.I., et al., *Cdc13p: a single-strand telomeric DNA-binding protein with a dual role in yeast telomere maintenance.* Science, 1996. 274(5285): p. 249-52.
76. Weinert, T.A. and L.H. Hartwell, *Cell cycle arrest of cdc mutants and specificity of the RAD9 checkpoint.* Genetics, 1993. 134(1): p. 63-80.
77. Garvik, B., M. Carson, and L. Hartwell, *Single-stranded DNA arising at telomeres in cdc13 mutants may constitute a specific signal for the RAD9 checkpoint.* Mol Cell Biol, 1995. 15(11): p. 6128-38.
78. Baumann, P. and T.R. Cech, *Pot1, the putative telomere end-binding protein in fission yeast and humans.* Science, 2001. 292(5519): p. 1171-5.

79. Baumann, P., E. Podell, and T.R. Cech, *Human Pot1 (protection of telomeres) protein: cytolocalization, gene structure, and alternative splicing.* Mol Cell Biol, 2002. 22(22): p. 8079-87.
80. Tani, A. and M. Murata, *Alternative splicing of Pot1 (Protection of telomere)-like genes in Arabidopsis thaliana.* Genes Genet Syst, 2005. 80(1): p. 41-8.
81. Wei, C. and C.M. Price, *Cell cycle localization, dimerization, and binding domain architecture of the telomere protein cPot1.* Mol Cell Biol, 2004. 24(5): p. 2091-102.
82. Wu, L., et al., *Pot1 deficiency initiates DNA damage checkpoint activation and aberrant homologous recombination at telomeres.* Cell, 2006. 126(1): p. 49-62.
83. Hockemeyer, D., et al., *Recent expansion of the telomeric complex in rodents: Two distinct POT1 proteins protect mouse telomeres.* Cell, 2006. 126(1): p. 63-77.
84. Lei, M., P. Baumann, and T.R. Cech, *Cooperative binding of single-stranded telomeric DNA by the Pot1 protein of Schizosaccharomyces pombe.* Biochemistry, 2002. 41(49): p. 14560-8.
85. Loayza, D. and T. De Lange, *POT1 as a terminal transducer of TRF1 telomere length control.* Nature, 2003. 423(6943): p. 1013-8.
86. Colgin, L.M., et al., *Human POT1 facilitates telomere elongation by telomerase.* Curr Biol, 2003. 13(11): p. 942-6.
87. Armbruster, B.N., et al., *Rescue of an hTERT mutant defective in telomere elongation by fusion with hPot1.* Mol Cell Biol, 2004. 24(8): p. 3552-61.
88. Yang, Q., Y.L. Zheng, and C.C. Harris, *POT1 and TRF2 cooperate to maintain telomeric integrity.* Mol Cell Biol, 2005. 25(3): p. 1070-80.
89. Horvath, M.P., et al., *Crystal structure of the Oxytricha nova telomere end binding protein complexed with single strand DNA.* Cell, 1998. 95(7): p. 963-74.
90. Hughes, T.R., et al., *The Est3 protein is a subunit of yeast telomerase.* Curr Biol, 2000. 10(13): p. 809-12.
91. Mitton-Fry, R.M., et al., *Structural basis for telomeric single-stranded DNA recognition by yeast Cdc13.* J Mol Biol, 2004. 338(2): p. 241-55.
92. Lei, M., et al., *DNA self-recognition in the structure of Pot1 bound to telomeric single-stranded DNA.* Nature, 2003. 426(6963): p. 198-203.
93. Lei, M., E.R. Podell, and T.R. Cech, *Structure of human POT1 bound to telomeric single-stranded DNA provides a model for chromosome end-protection.* Nat Struct Mol Biol, 2004. 11(12): p. 1223-9.
94. Murzin, A.G., *OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences.* EMBO J, 1993. 12(3): p. 861-7.
95. Froelich-Ammon, S.J., et al., *Modulation of telomerase activity by telomere DNA-binding proteins in Oxytricha.* Genes Dev, 1998. 12(10): p. 1504-14.
96. Houghtaling, B.R., et al., *A dynamic molecular link between the telomere length regulator TRF1 and the chromosome end protector TRF2.* Curr Biol, 2004. 14(18): p. 1621-31.
97. Liu, D., et al., *PTOP interacts with POT1 and regulates its localization to telomeres.* Nat Cell Biol, 2004. 6(7): p. 673-80.

98. Ye, J.Z., et al., *POT1-interacting protein PIP1: a telomere length regulator that recruits POT1 to the TIN2/TRF1 complex.* Genes Dev, 2004. 18(14): p. 1649-54.
99. Colgin, L. and R. Reddel, *Telomere biology: a new player in the end zone.* Curr Biol, 2004. 14(20): p. R901-2.
100. Fang, G. and T.R. Cech, *Oxytricha telomere-binding protein: DNA-dependent dimerization of the alpha and beta subunits.* Proc Natl Acad Sci U S A, 1993. 90(13): p. 6056-60.
101. Friedman, A.D., *Runx1, c-Myb, and C/EBPalpha couple differentiation to proliferation or growth arrest during hematopoiesis.* J Cell Biochem, 2002. 86(4): p. 624-9.
102. Konig, P., et al., *The crystal structure of the DNA-binding domain of yeast RAP1 in complex with telomeric DNA.* Cell, 1996. 85(1): p. 125-36.
103. Ogata, K., et al., *Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices.* Cell, 1994. 79(4): p. 639-48.
104. Court, R., et al., *How the human telomeric proteins TRF1 and TRF2 recognize telomeric DNA: a view from high-resolution crystal structures.* EMBO Rep, 2005. 6(1): p. 39-45.
105. Bilaud, T., et al., *The telobox, a Myb-related telomeric DNA binding motif found in proteins from yeast, plants and human.* Nucleic Acids Res, 1996. 24(7): p. 1294-303.
106. Vassetzky, N.S., et al., *Taz1p and Teb1p, two telobox proteins in Schizosaccharomyces pombe, recognize different telomere-related DNA sequences.* Nucleic Acids Res, 1999. 27(24): p. 4687-94.
107. Conrad, M.N., et al., *RAP1 protein interacts with yeast telomeres in vivo: overproduction alters telomere structure and decreases chromosome stability.* Cell, 1990. 63(4): p. 739-50.
108. Lustig, A.J., S. Kurtz, and D. Shore, *Involvement of the silencer and UAS binding protein RAP1 in regulation of telomere length.* Science, 1990. 250(4980): p. 549-53.
109. Buck, S.W. and D. Shore, *Action of a RAP1 carboxy-terminal silencing domain reveals an underlying competition between HMR and telomeres in yeast.* Genes Dev, 1995. 9(3): p. 370-84.
110. Longtine, M.S., et al., *A yeast telomere binding activity binds to two related telomere sequence motifs and is indistinguishable from RAP1.* Curr Genet, 1989. 16(4): p. 225-39.
111. Shore, D., *RAP1: a protean regulator in yeast.* Trends Genet, 1994. 10(11): p. 408-12.
112. Krauskopf, A. and E.H. Blackburn, *Control of telomere growth by interactions of RAP1 with the most distal telomeric repeats.* Nature, 1996. 383(6598): p. 354-7.
113. Marcand, S., E. Gilson, and D. Shore, *A protein-counting mechanism for telomere length regulation in yeast.* Science, 1997. 275(5302): p. 986-90.
114. Zhong, Z., et al., *A mammalian factor that binds telomeric TTAGGG repeats in vitro.* Mol Cell Biol, 1992. 12(11): p. 4834-43.
115. Chong, L., et al., *A human telomeric protein.* Science, 1995. 270(5242): p. 1663-7.

116. Bianchi, A., et al., *TRF1 binds a bipartite telomeric site with extreme spatial flexibility.* EMBO J, 1999. 18(20): p. 5735-44.
117. Griffith, J., A. Bianchi, and T. de Lange, *TRF1 promotes parallel pairing of telomeric tracts in vitro.* J Mol Biol, 1998. 278(1): p. 79-88.
118. Fairall, L., et al., *Structure of the TRFH dimerization domain of the human telomeric proteins TRF1 and TRF2.* Mol Cell, 2001. 8(2): p. 351-61.
119. van Steensel, B. and T. de Lange, *Control of telomere length by the human telomeric protein TRF1.* Nature, 1997. 385(6618): p. 740-3.
120. Bilaud, T., et al., *Telomeric localization of TRF2, a novel human telobox protein.* Nat Genet, 1997. 17(2): p. 236-9.
121. Broccoli, D., et al., *Comparison of the human and mouse genes encoding the telomeric protein, TRF1: chromosomal localization, expression and conserved protein domains.* Hum Mol Genet, 1997. 6(1): p. 69-76.
122. Broccoli, D., et al., *Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2.* Nat Genet, 1997. 17(2): p. 231-5.
123. Chen, Y., et al., *A shared docking motif in TRF1 and TRF2 used for differential recruitment of telomeric proteins.* Science, 2008. 319(5866): p. 1092-6.
124. Smogorzewska, A., et al., *Control of human telomere length by TRF1 and TRF2.* Mol Cell Biol, 2000. 20(5): p. 1659-68.
125. Ancelin, K., C. Brun, and E. Gilson, *Role of the telomeric DNA-binding protein TRF2 in the stability of human chromosome ends.* Bioessays, 1998. 20(11): p. 879-83.
126. Cooper, J.P., et al., *Regulation of telomere length and function by a Myb-domain protein in fission yeast.* Nature, 1997. 385(6618): p. 744-7.
127. Li, B., S. Oestreich, and T. de Lange, *Identification of human Rap1: implications for telomere evolution.* Cell, 2000. 101(5): p. 471-83.
128. Tomita, K., et al., *Competition between the Rad50 complex and the Ku heterodimer reveals a role for Exo1 in processing double-strand breaks but not telomeres.* Mol Cell Biol, 2003. 23(15): p. 5186-97.
129. Miller, K.M., M.G. Ferreira, and J.P. Cooper, *Taz1, Rap1 and Rif1 act both interdependently and independently to maintain telomeres.* EMBO J, 2005. 24(17): p. 3128-35.
130. Nakamura, T.M., J.P. Cooper, and T.R. Cech, *Two modes of survival of fission yeast without telomerase.* Science, 1998. 282(5388): p. 493-6.
131. Teixeira, M.T., et al., *Telomere length homeostasis is achieved via a switch between telomerase- extendible and -nonextendible states.* Cell, 2004. 117(3): p. 323-35.
132. Kanoh, J. and F. Ishikawa, *spRap1 and spRif1, recruited to telomeres by Taz1, are essential for telomere function in fission yeast.* Curr Biol, 2001. 11(20): p. 1624-30.
133. Chikashige, Y. and Y. Hiraoka, *Telomere binding of the Rap1 protein is required for meiosis in fission yeast.* Curr Biol, 2001. 11(20): p. 1618-23.
134. Spink, K.G., R.J. Evans, and A. Chambers, *Sequence-specific binding of Taz1p dimers to fission yeast telomeric DNA.* Nucleic Acids Res, 2000. 28(2): p. 527-33.
135. Li, B., A. Espinal, and G.A. Cross, *Trypanosome telomeres are protected by a homologue of mammalian TRF2.* Mol Cell Biol, 2005. 25(12): p. 5011-21.

136. Liu, D., et al., *Telosome, a mammalian telomere-associated complex formed by multiple telomeric proteins.* J Biol Chem, 2004. 279(49): p. 51338-42.
137. O'Connor, M.S., et al., *The human Rap1 protein complex and modulation of telomere length.* J Biol Chem, 2004. 279(27): p. 28585-91.
138. Ye, J.Z., et al., *TIN2 binds TRF1 and TRF2 simultaneously and stabilizes the TRF2 complex on telomeres.* J Biol Chem, 2004. 279(45): p. 47264-71.
139. de Lange, T., *Shelterin: the protein complex that shapes and safeguards human telomeres.* Genes Dev, 2005. 19(18): p. 2100-10.
140. Sarthy, J., et al., *Human RAP1 inhibits non-homologous end joining at telomeres.* EMBO J, 2009.
141. Miyoshi, T., et al., *Fission yeast Pot1-Tpp1 protects telomeres and regulates telomere length.* Science, 2008. 320(5881): p. 1341-4.
142. Bianchi, A. and D. Shore, *How telomerase reaches its end: mechanism of telomerase regulation by the telomeric complex.* Mol Cell, 2008. 31(2): p. 153-65.
143. Griffith, J.D., et al., *Mammalian telomeres end in a large duplex loop.* Cell, 1999. 97(4): p. 503-14.
144. Stansel, R.M., T. de Lange, and J.D. Griffith, *T-loop assembly in vitro involves binding of TRF2 near the 3' telomeric overhang.* EMBO J, 2001. 20(19): p. 5532-40.
145. Kim, N.W., et al., *Specific association of human telomerase activity with immortal cells and cancer.* Science, 1994. 266(5193): p. 2011-5.
146. Steinert, S., et al., *Telomere biology and cellular aging in nonhuman primate cells.* Exp Cell Res, 2002. 272(2): p. 146-52.
147. Cong, Y.S., W.E. Wright, and J.W. Shay, *Human telomerase and its regulation.* Microbiol Mol Biol Rev, 2002. 66(3): p. 407-25, table of contents.
148. Harley, C.B., A.B. Futcher, and C.W. Greider, *Telomeres shorten during ageing of human fibroblasts.* Nature, 1990. 345(6274): p. 458-60.
149. Bodnar, A.G., et al., *Extension of life-span by introduction of telomerase into normal human cells.* Science, 1998. 279(5349): p. 349-52.
150. Healy, K.C., *Telomere dynamics and telomerase activation in tumor progression: prospects for prognosis and therapy.* Oncol Res, 1995. 7(3-4): p. 121-30.
151. Harley, C.B. and B. Villeponteau, *Telomeres and telomerase in aging and cancer.* Curr Opin Genet Dev, 1995. 5(2): p. 249-55.
152. Sharma, H.W., et al., *Telomeres, telomerase and cancer: is the magic bullet real?* Anticancer Res, 1996. 16(1): p. 511-5.
153. Shay, J.W. and W.E. Wright, *Telomerase activity in human cancer.* Curr Opin Oncol, 1996. 8(1): p. 66-71.
154. Shay, J.W., H. Werbin, and W.E. Wright, *Telomeres and telomerase in human leukemias.* Leukemia, 1996. 10(8): p. 1255-61.
155. Harley, C.B. and S.W. Sherwood, *Telomerase, checkpoints and cancer.* Cancer Surv, 1997. 29: p. 263-84.
156. Deville, L., J. Hillion, and E. Segal-Bendirdjian, *Telomerase regulation in hematological cancers: a matter of stemness?* Biochim Biophys Acta, 2009. 1792(4): p. 229-39.
157. Hamilton, S.E. and D.R. Corey, *Telomerase: anti-cancer target or just a fascinating enzyme?* Chem Biol, 1996. 3(11): p. 863-7.

158. Shay, J.W., H. Werbin, and W.E. Wright, *Telomerase assays in the diagnosis and prognosis of cancer.* Ciba Found Symp, 1997. 211: p. 148-55; discussion 155-9.
159. Folini, M., P. Gandellini, and N. Zaffaroni, *Targeting the telosome: therapeutic implications.* Biochim Biophys Acta, 2009. 1792(4): p. 309-16.
160. Smith, C.D., et al., *Telomeric protein distributions and remodeling through the cell cycle in Saccharomyces cerevisiae.* Mol Biol Cell, 2003. 14(2): p. 556-70.
161. Ducrest, A.L., et al., *Regulation of the human telomerase reverse transcriptase gene.* Oncogene, 2002. 21(4): p. 541-52.
162. Lin, S.Y. and S.J. Elledge, *Multiple tumor suppressor pathways negatively regulate telomerase.* Cell, 2003. 113(7): p. 881-9.
163. Hahn, W.C., et al., *Enumeration of the simian virus 40 early region elements necessary for human cell transformation.* Mol Cell Biol, 2002. 22(7): p. 2111-23.
164. Hahn, W.C., et al., *Creation of human tumour cells with defined genetic elements.* Nature, 1999. 400(6743): p. 464-8.
165. Bryan, T.M. and R.R. Reddel, *Telomere dynamics and telomerase activity in in vitro immortalised human cells.* Eur J Cancer, 1997. 33(5): p. 767-73.

CHAPTER 2

FUNCTIONAL AND STRUCTURAL ANALYSES OF POT1-TPP1 COMPLEX IN SINGLE-STRANDED TELOMERIC REGION

POT1 proteins have been identified from many other organisms as the homologues of the α subunit of *O. nova* TEBP. It provides the most widespread solution to chromosome end protection in eukaryotes [1-8]. However, whether higher eukaryotes have a TEBPβ homologue remains unknown. TPP1 is a telomere protein that simultaneously interacts with both POT1 and TIN2 [9-11] (Fig. 1.8). In this chapter, I first focus on the functional and structural characterization of the POT1-TPP1 complex (Section 2.1). In Section 2.2, I extend the discussion to the studies of a budding yeast telomerase protein Est3 and its functional and structural connection to mammalian TPP1. The results of Section 2.1 have been published in *Nature* (2007. **445**(7127): p. 506-10) and part of the results in Section 2.2 have been published in *Nat Struct Mol Biol* (2008. **15**(9): p. 985-9) as a co-author paper with our collaborator Dr. Neal Lue at Weill Medical College of Cornell Universtiy.

2.1 The POT1-TPP1 Telomere Complex is a Telomerase Processivity Factor

2.1.1 Abstract

Telomeres were originally defined as chromosome caps that prevent the natural ends of linear chromosomes from undergoing deleterious degradation and fusion events. POT1

(protection of telomeres) protein binds the single-stranded G-rich DNA overhangs at human chromosome ends and suppresses unwanted DNA repair activities. TPP1 is a previously identified binding partner of POT1 that has been proposed to form part of a six-protein shelterin complex at telomeres. Here, the crystal structure of a domain of human TPP1 reveals an oligonucleotide/oligosaccharide-binding fold that is structurally similar to the β-subunit of the telomere end-binding protein of a ciliated protozoan, suggesting that TPP1 is the missing β-subunit of human POT1 protein. Telomeric DNA end-binding proteins have generally been found to inhibit rather than stimulate the action of the chromosome end-replicating enzyme, telomerase. In contrast, we find that TPP1 and POT1 form a complex with telomeric DNA that increases the activity and processivity of the human telomerase core enzyme. We propose that POT1–TPP1 switches from inhibiting telomerase access to the telomere, as a component of shelterin, to serving as a processivity factor for telomerase during telomere extension.

2.1.2 Introduction

Telomeres, the specialized DNA-protein complexes found at the termini of all linear eukaryotic chromosomes, protect chromosomes from degradation and end-to-end fusion [12]. Telomeric DNA typically consists of tandem repeats of a short G-rich sequence oriented 5’ to 3’ towards the chromosome terminus, with the G-rich strand extending beyond its complement to form a 3’ overhang. In most eukaryotes, telomere length is maintained by telomerase, a specialized reverse transcriptase that adds telomeric DNA to the 3’ ends of chromosomes to ensure complete genome replication [13]. Telomerase is strongly upregulated in most cancer cells and has been studied as a plausible anti-caner

target [14].

A six-protein complex is though to protect the telomeres of human chromosomes. TRF1 and TRF2 directly bind double-stranded telomeric DNA [15, 16]. POT1 directly binds the single-stranded 3' extension at the chromosome end [1, 2, 5, 6, 8, 17], and these are bridged through protein-protein interactions involving TIN2 and TPP1 (refs [9-11, 18-22]). The sixth protein, RAP1, binds mostly to TRF2 (refs [21, 23]). Two functions have been proposed for this complex: protecting the natural chromosome end from being mistaken for a broken end and being subjected to DNA repair, and negative regulation of telomerase by sequestration of its telomeric DNA substrate. Both functions of this complex are captured by the name shelterin [24].

Because POT1-TPP1 has been viewed as a structural component of the telomere, we were surprised to find that it increases both the activity and processivity of core telomerase. This is the first protein complex shown to substantially activate telomerase processivity. The crystal structure of TPP1 shows high structural similarity to the β-subunit of TEBP (telomere end-binding protein) from *Oxytricha nova*, a ciliated protozoan [25-27]. Because POT1 is the human homologue of TEBPα [1, 2, 8, 28], it now appears that capping of telomeres by a TEBPα-β dimer is more conserved evolutionary than had been expected.

2.1.3 TPP1 and POT1 forms ternary complexes with ssDNA

Recombinant human TPP1 protein with an N-terminal deletion, TTP1 (90-544) (Fig. 2.1a), was overexpressed and purified from *Escherichia coli*. TPP1 (90-544) was chosen because the 87 N-terminal residues of TPP1 are functionally dispensable in human cells

[9, 11] and are not conserved among TPP1 proteins of different organisms [10, 11]. For simplicity, we hereafter use TPP1 to represent TPP1 (90-544) unless stated otherwise.

TPP1 was first identified as the binding protein of POT1 [10, 11]. It binds to the C-terminal of POT1. In order to clarify the protein-protein interaction between TPP1 and POT1, a serial of constructs were designed that cover different regions of TPP1. These proteins were expressed and purified in *E. coli* and their binding ability to POT-C was judged by whether they could be eluted out as a complex through a gel filtration column. The results indicate that the domain in TPP1 that is responsible for binding to POT1 is from amino acid 250 to 334. The detailed domain interaction in POT1 and TPP1 is shown in Fig. 2.1a.

POT1 is known as the single-stranded telomeric DNA binding protein. The detailed interaction between POT1 and human single-stranded telomeric DNA repeats has been studied thoroughly. However, how POT1-TPP1 complex binds to single-stranded telomeric DNA as well as the detailed kinetic information is yet unknown. In order to test this, an 18-nucleotide single-stranded telomeric DNA (primer a, $(TTAGGG)_3$) was incubated with increasing amounts of TPP1 with or without POT1, and binding was analyzed by electrophoretic mobility shift assay (EMSA). A previous study has already illustrated that primer contains two overlapped POT1 binding sites, as boxed in Fig. 2.2a. POT1 protein bound to the DNA, whereas TPP1 on its own did not, even at a high protein concentration (375nM) (Fig. 2.1b, lane 1-5). When TPP1 was added to the POT1-DNA mixture, however, an additional complex formed that migrated above the POT1-DNA complex (Fig. 2.1b, lane 6-11). By two criteria, this more slowly migrating complex contained TPP1. First, two size variants of TPP1, both of which contain POT1-

binding domain (Fig. 2.1a), produced electrophoretically distinct complexes (Fig. 2.1c). Second, addition of the anti-His antibody confirmed that the slower complex contained His-tagged TPP1 (Fig. 2.1d). The amount of DNA in the ternary complex was increased compared with the amount of DNA bound to POT1 alone (compare lanes 5 and 11 in Fig. 2.1b), suggesting higher affinity.

The equilibrium dissociation constants (K_d) of the protein complexes with various telomeric single-stranded (ss)DNAs were determined by gel shift assays. Primer a (Fig. 2.2a) bound POT1 with a K_d of 26 nM, and the stability of this complex was increased sixfold by addition of TPP1 (Fig. 2.3). Primers a5 and a3 (Fig. 2.2a) contain single-nucleotide substitutions that force POT1 to bind either to a 5' site (corresponding to an internal site on a long telomeric G-overhang) or to a 3'-proximal site (corresponding to end-capping) [29]. The POT1-TPP1 complex showed a substantial preference for 3'-end binding (K_d =0.7 versus 7.4 nM; compare circles in Fig. 2.2b, c). Notably, the dissociation constant of POT1-a5 is also tenfold higher than that of POT1-a3 (K_d =8.3 versus 89 nM) (compare triangles in Fig. 2.2b, c), suggesting that the 3' end preference of the POT1-TPP1 complex is mainly dictated by POT1. Measurements of the kinetic stabilities of the various complexes confirmed that the off rate of the POT1-TPP1 complex was decreased by addition of equimolar TPP1 (Fig. 2.3). Furthermore, the complexes with a3 were kinetically about tenfold more stable than those with a5 (Fig. 2.3). Taken together, these data indicate two POT1-TPP1 binding modes on telomeric DNA, a lower affinity one at internal sequences and a higher affinity one at the 3' end.

2.1.4 Structural conservation between TPP1 and TEBPβ

Functional and structural studies have established that POT1 is the human homologue of the *O. nova* TEBP α-subunit [1, 2, 8, 28]. Although there has been no report of a TEBP β-subunit in any organism besides *O. nova* and a related ciliate, *Stylonychia mytilis*[30], the DNA-binding properties of POT1-TPP1 closely resembled those of *O. nova* TEBPα-β [31], consistent with TPP1 being the human homologue of TEBPβ. In addition, their domain organizations reveled clear similarities (Fig. 2.1a). First, both TPP1 and TEBPβ use a central region (PBD in TPP1 and αBD in TEBPβ) to interact with the carboxyl-terminal domains of their binding partners (POT1-C and TEBPα-C). Second, primary sequence analysis of the N-terminal domain of TPP1 (residues 90-250) predicted a secondary structure pattern of α-β-β-β-α-β-β-α, where the bold region is characteristic of oligonucleotide/oligosaccharide-binding (OB) folds found in many telomere-binding proteins including TEBPβ [8, 26]. Third, the C-terminal domains of both TPP1 and TEBPβ (TPP1-C and TEBP-C) are not involved in the interaction with the α-subunits (POT1 and TEBPα) and have evolved to have distinct functions [9, 10, 21, 31, 32] (Fig. 2.1a).

For crystallization studies, we first purified the N-terminal fragment TPP-N (Fig. 2.1a), which can interact with POT1 and would correspond to the TEBPβ fragment in the TEBPα-β-ssDNA crystal structure [26]. However, TPP1-N was unstable by itself. Limited proteolysis and matrix-assisted laser desorption/ionization (MALDI) mass spectrometry identified a protease-resistant core domain of TPP1-N containing residues 89-260 (Fig. 2.4a); this domain corresponds to the predicted OB fold (Fig. 2.1a). Recombinant TPP1 (90-250) expressed from *E. coli* was crystallized (Fig. 2.4b), and the

structure was solved by single anomalous dispersion (SAD) and refined to a resolution of 2.7 Å (Table 2.1). The final model contains residues 90-243 (Fig. 2.5a).

The structure of TPP1 (90-250) reveals a typical OB-fold architecture comprising a highly curved five-stranded β-barrel [33, 34](Fig. 2.5a). Hereafter, we will refer to TPP1 (90-250) as TPP1-OB (Fig.). An unbiased search for structurally homologous proteins using Dali [35] revealed that the structure of TPP1-OB is most similar to that of the OB fold of the *O. nova* TEBP β-subunit [26]. The two structures can be superimposed with a root-mean-square deviation of 2.0 Å in the positions of 144 equivalent Cα atoms (Fig. 2.5b). Notably, this structurally conserved region includes not only the central β-barrel, but also three peripheral α-helices, suggesting that TPP1 and TEBPβ are homologous proteins (Fig. 2.5b); other OB folds, such as that of TEBPα, are more distantly related (Fig. 2.5c). In addition to the overall structural similarity, the OB folds of TPP1 and TEBPβ share several unique features. First, the loop connecting β5 and αC (L_{5C}), unlike in the OB folds of POT1 and TEBPα, adopts an extended conformation and packs across one side of the β-barrel, forcing helix αC to cap the bottom the barrel (Fig. 2.5b). Second, helix αB is in a modified position rotated almost 90° relative to the orientation normally observed in OB folds. Taken together, these structural similarities strongly support the notion that TPP1 is the human homologue of *O. nova* TEBPβ. Given that TPP1 has been identified in many other eukaryotes [10, 11], TPP1/TEBPβ may be an evolutionarily conserved telomere protein.

Despite the high degree of structural conservation, the sequences of OB folds of TPP1 and TEBPβ are markedly divergent and share only 11% identity (Fig. 2.6). Significant sequence and structural variation is particularly evident in the connecting loop

regions. TPP1-OB has a very long loop (20 rediues), L_{12}, between strands β1 and β2, which packs on helix αA and covers one end of the barrel (Fig. 2.5a, b). In contrast, strands β1 and β2 of TEBPβ are connected by a short two-residue turn (Fig. 2.5b). These marked variances in the loop regions explain the failure to detect the similarity between these OB folds by bioinformatics.

2.1.5 POT1-TPP1 is a telomerase processivity factor

We investigated the ability of telomerase to extended the POT1-TPP1-ssDNA ternary complex, expecting some inhibition consistent with the shelterin model [24]. Human core telomerase was reconstituted *in vitro* and immunopurified via the haemagglutinin (HA) tag on the telomerase catalytic subunit (TERT). Primer a5 has a single-nucleotide mutation that forces POT1 to bind to its 5' end (Fig. 2.2a), leaving a telomerase – extendible 3' tail [29]. Addition of POT1 and TPP1 to primer a5 markedly increased the telomerase product size distribution. Primer a5 was extended via more than 30 cycles of template copying (Fig. 2.7a, lane 4), whereas in the absence of the POT1-TPP1 complex, the first three cycles accounted for most of the extension (lane 1). Under conditions of vast primer excess, longer extension products result from processive extension, not from rebinding of previously extended products [36]. We confirmed that this condition still pertained in the case of the ternary complexes by showing that the extension was independent of concentration over a 2,000-fold range (Fig. 2.8).

These data emphasize the longer extension products, because ^{32}P-GMP incorporation increases with product size. Quantification showed that POT1-TPP1 provided a threefold increase in activity (Fig. 2.7b) and, after dividing each product by

the amount of GMP incorporated, a fourfold increase in processivity (Fig. 2.7c). $R_{1/2}$, the number of repeats synthesized before half of the chains have dissociated, increased from 0.78 repeats with DNA primer a5 to 3.3 repeats with the POT1-TPP1-DNA complex. Because this fourfold increase in processivity is cumulative, it has a very large effect on the production of longer products (see double-headed arrows in Fig. 2.7c-e). POT1 by itself produced a more modest stimulation of processivity, consistent with earlier results [29], and TPP1 by itself had a similar effect (Fig. 2.7c-e).

The increase in telomerase processivity with the POT1-TPP1 complex was unexpected, and we were concerned that it might be due to a small molecule accompanying the protein. However, when we fractionated TPP1 by gel filtration chromatography, the processivity activity clearly co-migrated with the main protein peak (Fig. 2.9), indicating that the enhanced processivity was TPP1-specific.

The effects of POT1 and TPP1 were even more dramatic with other DNA primers. Primer a3 binds POT1 mainly at its 3' end (Fig. 2.2a), as does primer b ($(GGTTAG)_3$) [29]. As expected, the addition of POT1 almost completely blocked the extension of these primers (Fig. 2.7a, lanes 6 and 10). When TPP1 was also added, however, telomere extension was rescued and processivity was increased five- to sixfold (Fig. 2.7a, lanes 8 and 12; also see Fig. 2.7d, e). To test the possibility that TPP1 induced sliding of POT1 from the preferred 3' end binding site to the 5' site, we mapped the position of the leading edge of the POT1-TPP1 complex on the DNA using snake-venom phosphodiesterase I, which degrades ssDNA exonucleolytically from the 3' end. Complexes of primer a3 or primer a with the POT1-TPP1 heterodimer showed that the protein mostly occupied the DNA 3' end, with a local change in accessibility relative to

the POT1-DNA complexes where the 3’ end of the DNA protrudes from the protein (Fig. 2.10). Although TPP1 did not relocate the bulk of POT1 to an internal site on the DNA, we propose that the proteins are in fact binding internally in a minority of the complexes [29]; the resulting 3’ overhang is then extended by telomerase with the enhanced processivity characteristic of POT1-TPP1 complexes.

Primer AGGG-a ($AGGG(TTAGGG)_3$) has four blocks of GGG and therefore folds into Hoogsteen base-paired G-quadruplex structures, which are inhibitory for telomerase extension (Fig. 2.7a, lane 13) [37]. POT1 can trap an open form of this DNA, allowing telomerase extension (lane 4). TPP1 by itself (lane 15) gave the typical twofold incrase in telomerase activity seen with all the primers (Fig. 2.7b), but telomerase still stalled after every nucleotide added to the G-quadtrplex. POT1-TPP1 again stimulated highly processive extension by telomerase (lane 16).

We next asked whether the enhanced telomerase processivity and activity were dependent on the POT1-TPP1 interaction by using a panel of POT1 and TPP1 deletion mutant proteins. POT1-N, which lacks the TPP1 interaction domain, and TPP1-OB, which lacks the POT1 interaction domain, both failed to endow telomerase with increased processivity and activity (Fig. 2.11). Only when both proteins had intact interaction domains was the processivity greatly stimulated (Fig. 2.11), confirming the important role of the POT1-TPP1 interaction in this activity. The purified POT1 interaction domain of TPP1, TPP1-PBD, was insufficient to activate telomerase in the presence of POT1 (Fig. 2.12).

2.1.6 Discussion

E. coli DNA polymerase achieves high processivity by means of an accessory protein that serves as a “sliding clamp’, encircling the DNA and preventing dissociation [38]. By analogy, the POT1-TPP1 complex might move with telomerase, binding the DNA just upstream from its 3’ end and inhibiting dissociation. Given the DNA sequence-specificity of POT1-TPP1 binding, the protein would not slide continuously along the DNA but would ratchet in 6- or 12-nucleotide steps. On the other hand, ssDNA is intrinsically much more flexible than double-stranded (ds)DNA. Thus, a clamp would not need to slide or ratchet to keep the ssDNA associated with telomerase, but could instead remain fixed while the newly synthesized telomeric repeats formed a larger and larger protruding loop. Another question for future research is whether a single POT1-TPP1 complex clamp is sufficient for increased processivity, or whether the protein must coat the elongating telomeric DNA.

In normal human somatic cells that lack telomerase, telomeres shrink by about 30-100 base pairs per replication cycle [39]. This provides a plausible estimate for the amount of DNA synthesized by telomerase at each chromosome end, and is similar to more direct measurements in yeast [40]. Whether such extension is processive or distributive *in vivo* is unknown. However, we note that the telomerase processivity achieved here in the presence of POT1-TPP1 is around four repeats or 24 nucleotides, which would mean that one or a few rounds of telomerase extension per cell cycle would be sufficient to maintain human telomeres.

Given that POT1 and TPP1 are components of a negative feedback loop of telomere length control [9-11, 17], we were surprised to find that the POT1-TPP1 complex can also function as a positive telomerase processivity factor. We propose a

three-state model of telomere length regulation that can reconcile the two apparently opposite functions of POT1-TPP1. (1) When POT1-TPP1 covers the 3’ terminus of the G-overhang, it sequesters the telomere and prevents binding of telomerase. (2) POT1-TPP1 is removed from its high-affinity 3’ binding site by an unidentified mechanism, which might, for example, involve post-translational modification and disruption of the shelterin complex. (3) The POT1-TPP1 complex then serves as a telomerase processivity factor during telomere extension. As the telomere is elongated and reaches a certain threshold, the newly synthesized repeats bind shelterin complexes, the 3’ end of the overhang is re-bound by POT1-TPP1, and further telomerase extension is inhibited (back to state 1). Further work will be needed to understand how switching between such telomere and telomerase complexes is achieved and regulated *in vivo.*

2.1.7 Methods and materials

Oligonucleotide preparation

Telomeric ssDNAs used in the DNA-binding assay were purchased from IDT and Invirogen and 5’ -end- labeled using polynucleotide kinase.

Protein expression and purification

Human TPP1 was expressed in *E. coli* BL21(DE3) using a modified pET28b vector with a SUMO protein fused at the N-terminus after the 6XHis tag. After induction for 16 hours with 0.1 mM IPTG at 25°C, the cells were harvested by centrifugation and the pellets were resuspended in lysis buffer (50 mM Tris-HCl pH 8.0, 50 mM NaH_2PO_4, 400 mM NaCl, 3 mM imidazole, 10% glycerol, 1 mM PMSF, 0.1 mg/ml lysozyme, 2 mM 2-

mercaptoethanol, and home-made protease inhibitor cocktail). The cells were then lysed by sonication and the cell debris was removed by ultracentrifugation. The supernatant was mixed with Ni-NTA agarose beads (Qiagen) and rocked for 6 hours at 4˚C before elution with 250 mM imidazole. Then ULP1 protease was added to remove the His_6-SUMO tag. The protein was further purified by passage through Mono-Q and by gel-filtration chromatography on Superdex200 equilibrated with 25 mM Tris-HCl pH 8.0, 150 mM NaCl. The purified protein was concentrated to 2-10 mg/ml and stored at -80˚C. The TPP1 deletion mutants (TPP1-OB, TPP1N, and TPP1-PBD) were cloned, expressed, and purified following the same procedure as described above. POT1 and its splicing variant 2, POT1N, which contains the entire DNA-binding domain, and TPP1-N (residue 89-334) were expressed by using baculovirus-insect cells and purified as described previously [8].

Limited protease (Glu-C) cleavage of TPP1N

TPP1N was incubated with 0.2% w/w Glu-C (Roche) at room temperature in 25 mM Tris-HCl pH 8.0, 150 mM NaCl, and 2 mM DTT. At various time points, 8 μl aliquots of the reaction were withdrawn, diluted with 12 μl of water and 5 μl of SDS loading dye, and run on 15% SDS–PAGE visualized with Coomassie brilliant blue stain.

MALDI mass spectrometry of the limited protease (Glu-C) cleavage products

For MALDI mass spectrometric analysis, TPP1N was incubated with 0.2% w/w Glu-C in 25 mM Tris-HCl pH 8.0, 150 mM NaCl, and 2 mM DTT at room temperature. Aliquots were withdrawn as described above for SDS–PAGE analysis. At the 60 min time point, 2

µl of the reaction mixture was co-crystallized with 2 µl sinapinic acid matrix. The samples were analyzed by MALDI-TOF-MS in linear mode. The major product by MALDI had an MH(+1) of 18,611 ± 30 Da. Examination of the map of predicted Glu-C sites revealed that this fragment corresponds to one predicted fragment: $TPP1_{90\text{-}260}$ [MH(+1) 18,597 Da].

Crystallization, data collection and structure determination of TPP1-OB

For crystallization, TPP1-OB was concentrated to 25 mg/ml in storage buffer. Crystals were grown by hanging drop vapor diffusion at 16ºC. The precipitant/well solution contained 100 mM trisodium citrate pH 5.6, 3.5 M sodium formate and 5 mM DTT. The crystals belong to space group $I4_122$ with unit cell dimensions of a = b = 117.25 Å, and c = 171.58 Å. The asymmetric unit contains a dimer of TPP1-OB. All crystals were gradually transferred into a harvesting solution containing 100 mM sodium citrate pH 5.6 and 5 M sodium formate, then flash-frozen in liquid nitrogen for storage and data collection under cryogenic conditions (100K). Data were collected at the Advanced Photon Source beamlines 23ID-D and processed using HKL2000 (Otwinowski and Minor, 1997). SAD data from the mercury derivative were used to obtain initial phases. Five mercury atoms were located and refined, and the SAD phases calculated using SHARP (E. d. La Fortelle, G. Bricogne, 1997); the initial SAD map was significantly improved by solvent flattening. A model was automatically built into the modified experimental electron density using ARP/WARP (V. S. Lamzin, A. Perrakis, K. S. Wilson, 2001); the model was then transferred into the native unit cell by rigid-body refinement and further refined using simulated-annealing and positional refinement in

CNS [41], with manual rebuilding (O) [42].

Electrophoresis Gel Mobility Shift Assay

Proteins in binding buffer (25 mM HEPES/NaOH, pH 7.5, 50 mM NaCl, 40 mM KCl, 7% glycerol, 1 mM EDTA, and 0.1 mM DTT) was mixed with 1 μM ^{32}P-labeled telomeric ssDNAs in a total volume of 20 μL. The reaction mixtures were incubated at room temperature for 30 min. Then the mixtures were directly loaded onto a 4-20% nondenaturing polyacrylamide gel. Electrophoresis was carried out in TBE buffer at 150 V for 85 min at 4 °C. The gels were dried, and radiolabeled ssDNA was visualized using a PhosphorImager.

***In vitro* Reconstitution of Human Telomerase**

C-terminal HA-tagged human TERT was expressed from phTERT-HA2 and hTER from phTR [29, 36] by using the TnT quick-coupled transcription/translation system (Promega). Each 500-μl reaction contained 400 μl of TnT quick mix, 40 μl of PCR enhancer (Promega), 20 μl of 1 mM methionine, 20 μl of water, and 1.05 μg of each plasmid DNA. After incubation at 30 °C for 2 h, the reconstituted telomerase complex was affinity-purified on anti-HA F7-agarose beads (Santa Cruz Biotechnology). Anti-HA F7-agarose beads (150 μl), washed with TMG-100, were added for immunoprecipitation at 4 °C overnight. The beads were washed with 1× telomerase assay buffer (see below) four times and then resuspended in 1× telomerase assay buffer. The quantity of ^{35}S-hTERT was determined.

Telomerase Activity Assay

Two types of telomerase activity assays, direct assay and telomeric repeat amplification protocol (TRAP) assay, have been developed and widely used in the studies of telomerase. In the direct assay, ^{32}P-radioactive labeled nucleotides (usually dGTP) is incorporated into the 3' end of the telomeric DNA primer. After the reaction is finished, the elongation products can be visualized in a denaturing polyacrylamide gel by autoradiograph [43]. When different telomeric primers with variant ending sequence are provided, the product's pattern on the gel will change accordingly. Similar results will also appear if the RNA template sequence is mutated. Therefore, the direct assay is a very powerful tool for studying how telomerase behaves in various reaction systems. However, the direct assay requires abundant amount of telomerase materials with good processivity in order to observed the clear "ladder" pattern. For human tissue with very low amount of telomerase or mouse cells bearing low processivity telomerase, TRAP assay is more suitable. In the TRAP assay, telomerase first adds telomeric repeats to the primer during the reaction. Then the reaction products will be amplified through polymerase chain reaction (PCR) by using special designed primers [44, 45]. TRAP assay takes the advantage of the amplification power of PCR and magnifies the very low telomerase activity to a detectable level. Therefore, TRAP assay is ideal to deal with human tissue or tumor samples with really low levels of telomerase activity. Although TRAP assay results show the similar hexa-nucleotide "ladder" as the direct assay does, the banding pattern comes from the PCR amplification rather than from telomerase repeat addition.

In this chapter, I applied the direct assay to monitor the change of telomerase

activity and processivity to study the function of the POT1-TPP1 complex. Activity of the immunopurified human telomerase complex reconstituted in vitro was determined by a direct assay modified from a published protocol [36]. Briefly, the reaction mixture (20 μl) contained 1× telomerase assay buffer (50 mm Tris-HCl, pH 8.0, 50 mm KCl, 1 mm $MgCl_2$, 5 mm 2-mercaptoethanol, and 1 mm spermidine), 100 nm telomeric DNA primer, 0.5 mm dATP, 0.5 mm dTTP, 2 μm dGTP, and 1.25 μm [α-^{32}P]dGTP (800 Ci/mmol) with 6 μl of immunopurified telomerase complex. The reaction was incubated at 30 °C for 1 h, and the products were precipitated with the addition of 100 μl of 3.6 m NH_4OAc, 20 μg of glycogen, and 450 μl of ethanol. After incubation at –80 °C for 1 h, samples were centrifuged at 4 °C for 20 min, and the pellets were washed with 70% ethanol and resuspended in 1× gel-loading buffer (40% formamide, 10 mm Tris-HCl, pH 8.0, 10 mm EDTA, and 0.05% xylene cyanol). The heat-denatured samples were loaded onto a 10% polyacrylamide, 1× TBE (Tris borate-EDTA), 7 M urea denaturing gel for electrophoresis. After electrophoresis, the gel was dried and quantified using a PhosphorImager.

Snake-venom phosphodiesterase I (SVPI) digestion

Protein-DNA complex was preformed before the addition of SVPI. The complex mixture (9 μl) contained 1.1x telomerase assay buffer (50 mM Tris-HCl pH 8.0, 50 mM KCl, 1 mM $MgCl_2$, 5 mM 2-mercaptoethanol and 1 mM spermidine), 0.11x protein buffer (25 mM Tris-HCl pH 8.0, 150 mM NaCl), protein or protein mixture (TPP1, POT1, or POT1-TPP1) and 5’ labeled ^{32}P-DNA and was incubated at room temperature for 30 min. Then 0.3 μg SVPI in 1 μl stock solution (100 mM Tris-Cl pH 8.9, 100 mM NaCl and 15 mM

$MgCl_2$) was added to start the digestion. The reaction was incubated at 30°C for 5 min and then stopped by the addition of 1 µl 100 mM EDTA. After heat-inactivation at 95°C for 2 min, 10 µl of 94% formamide, 1xTBE and loading dye were added to the sample. A portion (10 µl) of the final mixture was loaded onto a 10% polyacrylamide/1xTBE/7 M urea denaturing gel for electrophoresis.

2.2 Budding Yeast Est3 is a Putative Ortholog of Mammalian TPP1

2.2.1 Abstract

Ever shorter telomeres 3 (Est3) is an essential telomerase regulatory subunit thought to be unique to budding yeasts. Here we use multiple sequence alignment and hidden Markov model–hidden Markov model (HMM-HMM) comparison to uncover potential similarities between Est3 and the mammalian telomeric protein TPP1. Analysis of site-specific mutants of *Candida albicans* Est3 revealed functional distinctions between residues that are conserved between Est3 and TPP1 and those that are unique to Est3. Although both types of residues are important for telomere maintenance *in vivo*, only the former contributes to telomerase activity in vitro and facilitates the association of Est3 with telomerase core components. Consistent with a function in protein- protein interaction, the residues common to Est3 and Tpp1 map to one face of an OB-fold model structure, away from the canonical nucleic acid binding surface. We propose that Est3 and the OB-fold domain of TPP1 mediate a conserved function in telomerase regulation.

2.2.2 Introduction

Telomeres are specialized nucleoprotein structures that maintain the integrity of

eukaryotic chromosomal termini by protecting them from fusion and recombination, and promoting their replication (for reviews, see refs. [12, 13, 24, 46]). In most organisms, telomeric DNA consists of short repetitive sequences that are rich in G residues on the 3' end-containing strand. These repeats are maintained by a ribonucleo-protein (RNP) known as telomerase, which acts as an unusual reverse transcriptase (for reviews, see refs.[13, 47-49]). Both telomere binding proteins and telomerase are crucial for the maintenance of telomere integrity through multiple cell divisions, which in turn is pivotal in supporting genome stability and promoting cellular life span.

Remarkably, components of both the telomeric protein complex and the telomerase complex have been observed to be evolutionarily malleable. For example, the terminal G-strand overhangs of mammalian telomeres are bound by POT1 and TPP1, which are subunits of a larger complex known as shelterin. In contrast, the budding yeast telomere overhangs evidently interact with a Replication protein A–like heterotrimeric complex consisting of Cdc13, Stn1 and Ten1 [50, 51]. Although each of the mammalian and yeast proteins consists of a variable number of OB-fold domains, no orthologous relationship has been established between any protein pairs from either sequence or functional comparison. Similarly, no telomerase subunit except the catalytic protein (telomerase reverse transcriptase (TERT)) and template RNA (TER) is universally conserved. Instead, both mammals and yeast possess species-specific factors that are crucial for telomerase assembly and function. Thus, the fundamental importance of the telomere-maintenance machinery is in striking contrast to its evolutionary plasticity. Est3 is a small but functionally essential subunit of the yeast telomerase complex [52]. Deletion of *EST3* leads to a progressive telomere-attrition phenotype that mimics closely

deletion of TERT or TER [52, 53]. Yet the protein is not essential for the catalytic activity of telomerase *in vitro* [52, 54, 55]. Recent experiments, however, point to a role for *Candida albicans* Est3 in promoting holoenzyme assembly and stimulating the polymerization activity of telomerase in a primer substrate–dependent fashion [56]. To gain further insights into Est3 mechanisms, we used a combination of bioinformatic analysis and site-specific mutagenesis to identify structural attributes of the protein necessary for its biochemical functions. Our results reveal unexpected similarities between Est3 and mammalian TPP1.

2.2.3 Similarities between Est3 and TPP1

To delineate the structure and function of Est3, a multiple sequence alignment was first performed against 12 *EST3* homolog genes from different *Saccharomyces* and *Candida* species (5 from *Saccharomyces* and 7 from *Candida*) (Fig. 2.13a). The result was then used to query the HHpred server [57], which exploits comparison of hidden Markov models (HMMs) profiles for the identification of distantly related proteins. Notably, the top-scoring hit from this analysis was the mammalian telomeric protein TPP1 (E-value = 0.71, P-value = 91.5), indicating a possible connection between Est3 and TPP1. This result is consistent with our previous data that revealed an interaction between TPP1 and telomerase as well as a stimulatory effect of TPP1 on telomerase activity and processivity (Section 2.1).

2.2.4 Effects of CaEst3 mutations on telomeres, telomerase assembly and activity

To further illustrate the connection between Est3p and TPP1, we proceeded to generate a

series of site-specific *C. albicans* EST3 mutants designed on the basis of a composite sequence alignment between Est3 and TPP1 homologs (Fig. 2.13a). This sequence alignment highlights three residues that are well conserved in all Est3 and TPP1 family members (Trp36, Asp91 and Asp/Glu169 in *C. albicans* Est3 and Trp98, Asp148 and Glu215 in human TPP1), as well as residues that are conserved only in Est3 homologs (Arg72, Ser92, Phe109, Arg116, Thr121 and Gln170). These residues were replaced with alanine. The resulting mutants were reintegrated into an *C. albicans est3Δ/est3Δ* strain and subjected to functional analyses in Dr. Neal Lue's laboratory at the Weill Medical College of Cornell University.

Dr. Lue and colleagues first examined the impact of mutations on telomere lengths and contents by telomere Southern (Fig. 2.14). All of the mutants, with the exception of D59A, R72A and S92A, manifested substantial defects in telomere maintenance. Notably, both the D91A and D91A/S92A mutants behaved as null alleles with regard to telomere maintenance (Fig. 2.14a, b). Then, the levels of mutant Est3 proteins and their association with telomerase RNA (Ter1) were investigated by IgG-Sepharose pull-down, western blotting and reverse-transcription PCR (RT-PCR). All of the mutant proteins were present at near wild-type levels, except for D91A and DS91AA, which were reduced by approximately five-fold (Fig. 2.14c). On the other hand, quantitative RT-PCR revealed no detectable Ter1 in the D91A and DS91AA pull-down samples, and Ter1 levels were reduced to ~20% of the normal Ter1 levels in the W36A and D169A samples (Fig. 2.14d, e). Titration of the wild-type extract indicates that the absence of Ter1 in the D91A pull-down sample cannot be explained by reduced protein level. In particular, the D91A protein level is reduced approximately 5-fold, whereas the

associated Ter1 is reduced more than 25-fold (Fig. 2.14e). Thus, Asp91 enhances the stability of Est3 and is essential for telomerase association, whereas Trp36 and Asp169 are not required for protein stability but promote telomerase association. The other mutated residues seem dispensable for both functions. Finally, primer-extension activity was measured using the IgG-Sepharose precipitates to determine the effect of EST3 mutations on telomerase activity in vitro (Fig. 2.15). Notably, mutations in the residues conserved only in Est3 generally had little effect on telomerase activity on either primer (Fig. 2.15b). Telomerase from the W36A and D169A mutants showed greatly reduced activity (Fig. 2.15b). For both mutants, the magnitude of reduction in activity (approximately 50-fold) was far greater than that of telomerase RNA (approximately 5-fold) (compare Figs. 2.14d and 2.15c), indicating a dramatic reduction in specific activity. Thus, Trp36 and Asp169 must have a function in stimulating telomerase activity that goes beyond their role in promoting telomerase association of Est3, possibly through an allosteric effect. As expected, no telomerase activity was detected in the D91A and DS91AA IgG-Sepharose precipitates as a result of their lack of telomerase association (Fig. 2.15b). (detailed data refer to ref. [58]Fig. 2 and 3)

2.2.5 Effects of TPP1 mutations on telomerase activity

We generated three TPP1N (amino acid 89-334) mutants in *E. coli*: W98A, D148A, and E215A, which are equivalent to Trp36, Asp91 and Asp/Glu169 in *C. albicans* Est3. Both D148A in TPP1 and D91A in Est3 resulted in insoluble proteins, suggesting that this Asp is important for protein stability. The other two mutant proteins as well as the wild type TPP1N were then purified and used in the *in vitro* telomerase activity assay. The addition

of POT1 to the wild type TPP1N and primer a5 gave us a typical increased telomerase product size distribution. As expected, primer a5 was extended via more than 30 cycles of template copying when both POT1 and TPP1N are present. (Fig. 2.16, lane 1-3) Notably, when W98A was introduced, it greatly abolished POT1-TPP1 complex's processive effect on telomerase for about 50 folds (Fig. 2.16, lane 4), which is similar as Est3 when mutating Trp36. However, the other mutant (E215A) didn't have such an effect. The DNA primer was extended just like the wild type did (Fig. 2.16, compare lane 3 with lane 5). So far, two of the mutations (Trp and Asp) at the well-conserved residues in Est3p and TPP1 have detectable phenotypes. In the contrast, the other mutants mutating at the sites only shared by Est3 family members just function similar to wild type (The data are summarized in Table 2.2). These results further revealed the functional connection between TPP1 and *Ca*Est3.

2.2.6 Studies towards the structure of Est3

The most direct approach to prove the Est3-TPP1 connection is to determine the three-dimensional structure of Est3. To this end, full-length Est3 proteins from several different budding yeast species including *S. cerevisiae, C. albicans,* and *C. tropicalis* were expressed, purified and used in extensive crystallization trials. However, we failed to obtain any crystals of these proteins. The full-length *S. cerevisiae* Est3 (*Sc*Est3) presents dual bands in SDS-PAGE, which suggest that it is not stable in solution. Limited proteolysis and mass spectrometry analyses identified a protease-resistant core of *Sc*Est3 (residues 2-181). *Sc*Est3$_{2\text{-}181}$ was then prepared and subjected to crystallization screening. A cluster of needle shaped tiny crystals showed up after more than 100 days under the

condition of HEPES pH 7.5 100 mM, PEG 4000 30%, NaCl 200 mM (Fig. 2.17a) at 4 °C. During the protein purification process, I noticed that there was a monomer-dimer equilibration for *Sc*Est3 in solution. I hypothesized that it was this equilibration process that caused the heterogeneity of the sample and the difficulties in crystallization. Further work is needed to solve this issue to obtain Est3 crystals suitable for structure determination.

As an alternative approach to X-ray crystallography, I also employed NMR spectrometry to study the solution structure of Est3 (with the help from Dr. Erik Zuiderweg at the University of Michigan). The ^{1}H-^{15}N 2D spectrum analysis for the ^{15}N-labeled *Candida. tropicalis* Est3 (residue 4-196) was showed in Fig. 2.17b. There are about 170 ^{1}H-^{15}N isolated peaks nicely distributed on the spectrum, suggesting that the *Ct*Est3 protein is well folded. The positions of these peaks also indicate a β-strand-rich structure. This is consistent with our prediction that *Ct*Est3 is an OB-fold containing TPP1-like protein. However, the *Ct*Est3 protein tends to precipitate after ~ 5 hours at 25° (at which NMR data were collected), preventing us from obtaining the 3D spectra for structure determination. Further work in needed to improve the thermal stability of *Ct*Est3. Alternatively, Est3 proteins from other budding yeast strains will also be used in the NMR study.

2.2.7 Discussion

Altogether, functional analysis of Est3 mutants has revealed three classes of important residues (class I to III in Table 2.2). Class I, represented by Asp91, is essential for both the *in vivo* and *in vitro* function of Est3. Substitution of Asp91 results in the concurrent

dissociation of Est3 and Est1 from the telomerase complex and primer-dependent loss of telomerase activity, thus leading to progressive telomere attrition. Class II residues, including Trp36 and Asp169, are less important functionally. Substitution of these residues results in partial dissociation of Est3 and Est1 from telomerase and partial loss of telomerase activity. The extent of telomere loss in these mutants is accordingly less severe than in the D91A mutant. Class III residues, including Phe109, Arg116, Thr121 and Gln170, are also required for normal telomere maintenance. However, mutations in these residues have little effect on the telomerase association of Est3 and telomerase activity. The existence of class III residues points to an additional molecular function for Est3 that is not revealed by the current *in vitro* assay.

Notably, class I and II residues are conserved between Est3 and Tpp1, whereas class III residues are shared only by Est3 family members. The crystal structure of TPP1-OB allows us to locate the three TPP1 residues that are equivalent to those in *Candida. albicans* Est3. All three residues (Trp98, Asp148 and Glu215 in TPP1, equivalent to Trp36, Asp91 and Asp169 in *C. albicans* Est3) stay at one face of the OB-fold, which is away from the classical oligomer-binding site (Fig. 2.18b). This suggests that these residues are not crucial for interacting with DNA but may be involved in protein-protein interactions. Asp148 (equivalent to Asp91 in *C. albicans* Est3), which is localized at the end of β2 in TPP1-OB, is able to form four hydrogen bounds with the backbone amino groups (W98, I99, T150 and H151) through its side chain. Thus, the helix αA at the end of the β2 of the OB-fold is stabilized. By analogy, this will explain the biological significance of Asp91 in *C. albicans* Est3. Remarkably, Asp91 is well conserved through many other OB-fold-containing proteins besides TPP1 and Est3. It seems that the

conserved Asp91 is important for the correct folding of the OB-fold. In contrast, class II residues (Trp36 and Asp169) are shared only by Est3 and TPP1 (Fig. 2.18a). Consistent with this notion, Lue's group has been unable to detect DNA binding by Est3 using various substrates. A plausible protein target of Est3 may be the N-terminal domain of TERT, which is conserved between yeast and human telomerase, and has been linked to Est3 genetically [59].

The structural and functional similarities uncovered in this study suggest that Est3 could be orthologous to the OB-fold domain of TPP1. This then begs the question as to how orthologous proteins can be components of different macromolecular complexes in different organisms. A speculative evolutionary scenario is as follows (Fig. 2.19): the ancestral TPP1/Est3 was a component of the telomeric protein complex and, as such, a stimulatory factor for telomerase action. It has been suggested that, during budding yeast evolution, POT1 might have been lost from telomeres as a result of mutations in the telomere repeat sequence [23], thus resulting in the concomitant dissociation of TPP1. The yeast mutant would be expected to face profound selection pressures for alternative mechanisms of telomere protection and telomerase stimulation. The telomere protective function was apparently assumed by the Cdc13–Stn1–Ten1 complex. The telomerase-stimulatory function, on the other hand, was rescued by new interactions between Est3 and the telomerase protein Est1, which stabilized the association between Est3 and TERT [56, 60]. In this scenario, Est3 is orthologous to only the telomerase-stimulatory domain of TPP1, having lost the region(s) required for telomere localization (for example, the TIN2 and POT1 binding domain). The fact that budding yeast has retained Est3 despite the drastic remodeling of its telomere-maintenance machinery suggests that this

telomerase-stimulatory function may be universally required. In this regard, it would be interesting to determine whether the ciliate TEBPβ and *S. pombe* Tpz1, which are homologous to TPP1 (refs. [61-64], also mediate telomerase stimulation. Moreover, an apparent homolog of Est3 and TPP1 has not been identified in plants. Instead, a POT1 homolog is reportedly associated with the telomerase RNP. Understanding the shared and distinctive properties of these factors would surely provide valuable insights on telomere evolution.

2.2.8 Methods and materials

Protein expression and purification

Human TPP1 was expressed in *E. coli* BL21(DE3) using a modified pET28b vector with a SUMO protein fused at the N-terminus after the 6XHis tag. After induction for 16 hours with 0.1 mM IPTG at 25°C, the cells were harvested by centrifugation and the pellets were resuspended in lysis buffer (50 mM Tris-HCl pH 8.0, 50 mM NaH_2PO_4, 400 mM NaCl, 3 mM imidazole, 10% glycerol, 1 mM PMSF, 0.1 mg/ml lysozyme, 2 mM 2-mercaptoethanol, and home-made protease inhibitor cocktail). The cells were then lysed by sonication and the cell debris was removed by ultracentrifugation. The supernatant was mixed with Ni-NTA agarose beads (Qiagen) and rocked for 6 hours at 4˚C before elution with 250 mM imidazole. Then ULP1 protease was added to remove the His_6-SUMO tag. The protein was further purified by passage through Mono-Q and by gel-filtration chromatography on Superdex200 equilibrated with 25 mM Tris-HCl pH 8.0, 150 mM NaCl. The purified protein was concentrated to 2-10 mg/ml and stored at -80˚C.

The Est3 proteins (*Sc*Est3, *Ca*Est3 and *Ct*Est3) were cloned into pET28b vector

with a SUMO protein fused at the N-terminus after the 6XHis tag. Proteins were expressed and purified following the same procedure as described above. The *Ct*Est3 protein for the NMR study, cells were allowed to grow overnight in 5 ml of LB media at 37 °C and subsequently transferred into 2 L of M9 minimal media containing $^{15}NH_4Cl$ (1.5 g/L) (Cambridge Isotope Laboratories, Inc.) for preparation of ^{15}N-labeled protein.

In vitro telomerase activity assay

Telomeric ssDNAs used in the DNA-binding assay were purchased from IDT. For *in vitro* telomerase activity assays [29], C-terminal HA-tagged HA-tagged human TERT was expressed from phTERT-HA2 and hTR from phTR [36] using the TnT quick-coupled transcription/translation system (Promega). And the reconstituted core telomerase was affinity-purified on anti-HA F7-agarose beads (Santa Cruz Biotechnology) [29].

Sequence alignment and structure prediction

The sequences of 12 Est3 homologues from budding yeast were culled from the National Center for Biotechnology Information (NCBI) and Broad Institute databases and aligned using T-coffee (http://tcoffee.vital-it.ch/cgi-bin/Tcoffee/tcoffee_cgi/index.cgi). The alignment was used to query the HHpred server, resulting in the identification of human TPP1 as a potential ortholog [57]. A composite alignment of the 12 Est3 and 3 TPP1 homologs (from humans, mice and *Xenopus laevis*) was then generated using the ProbCons server [65].

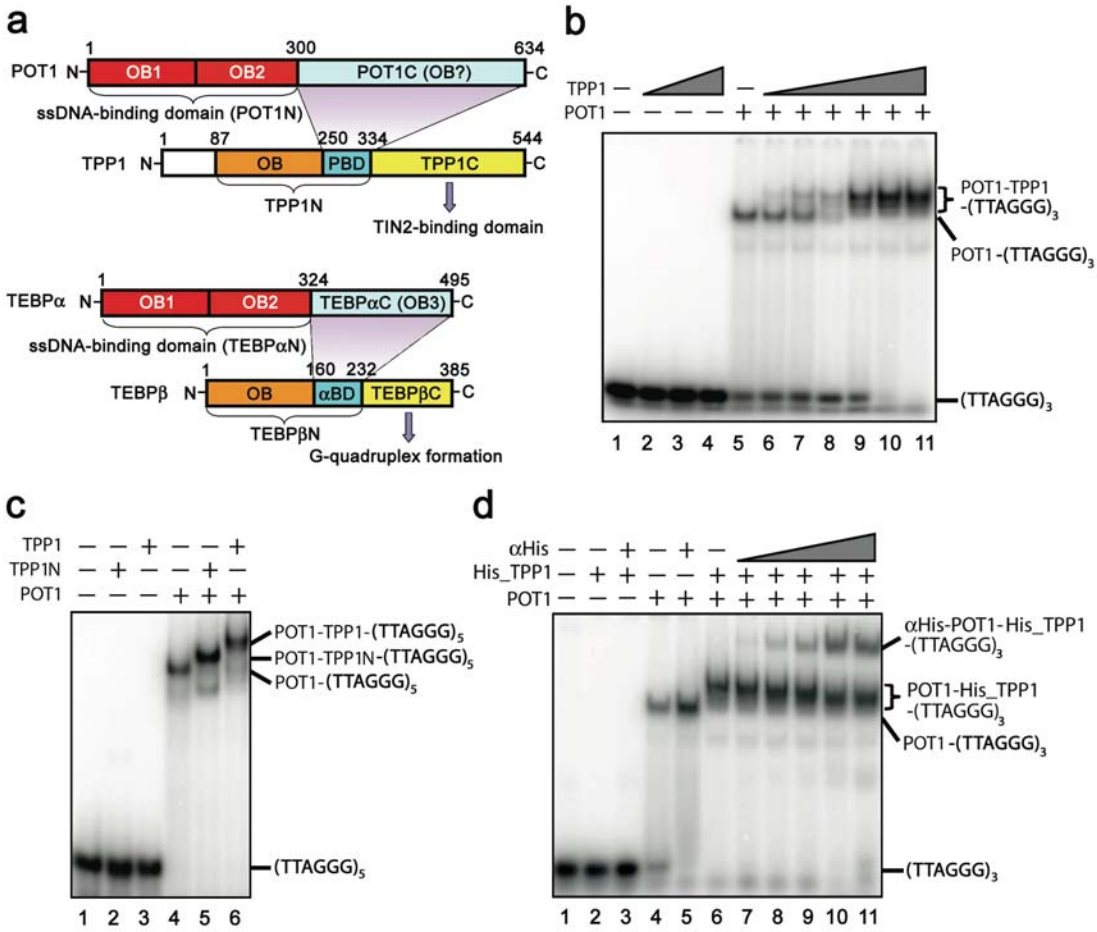


Figure 2.1 TPP1 binds to the POT1-ssDNA complex and enhances the POT1-ssDNA interaction.

a) Human POT1-TPP1 and *O. nova* TEBPα-β complexes share similar domain organization. In POT1 and TEBPα, the N-terminal ssDNA-binding domains are colored in red, and the C-terminal TPP1/TEBPβ-binding domains are in light blue. In TPP1 and TEBPβ, the N-terminal OB folds are in orange, the central α-subunit-binding region (PBD in TPP1 and αBD in TEBPβ) is in cyan, and the C-terminal domains are in yellow. Numbers indicated amino acid positions at the boundaries of various subdivisions.

b) TPP1 requires POT1 in order to interact stably with telomeric DNA. Primer a (50 nM) was incubated with increasing amounts of TPP1 (2, 10 and 50 nM in lanes 2-4, 2-100 nM in lane 6-11) in the absence or presence of POT1.

c) Both TPP1 and TPP1-N generate super-shifted species but with different mobility; use of a longer primer enhanced the separation.

d) The addition of anti-His antibody detected the existence of His-tagged TPP1 in the slower migrating complex.

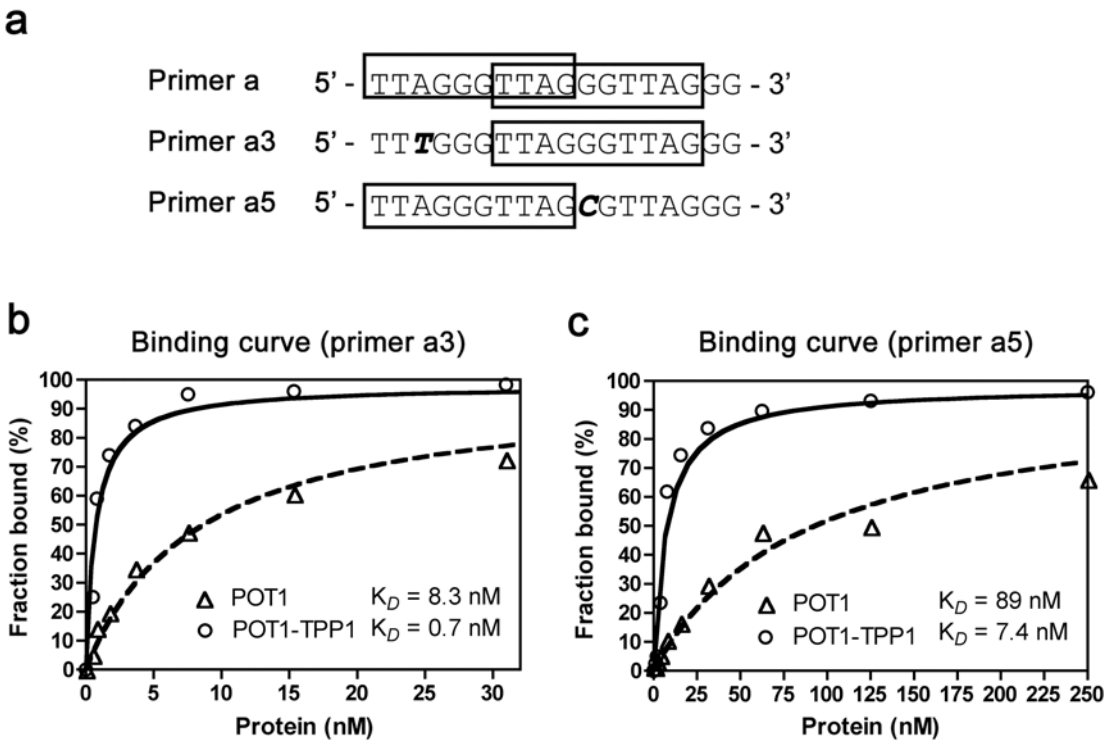
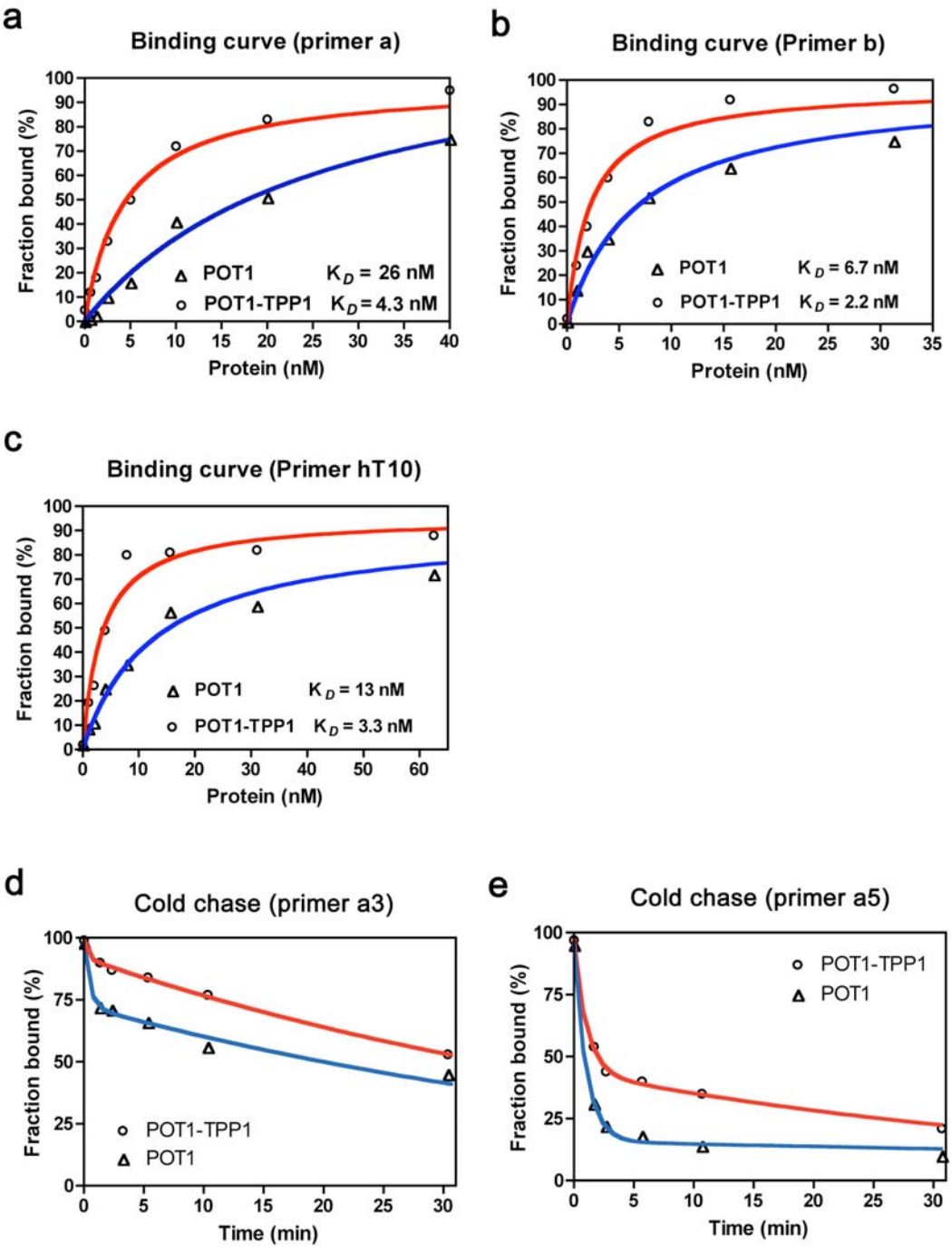


Figure 2.2 The POT1-TPP1 complex binds to the single-stranded telomeric overhang with 3' end preference.
(a) Sequence of primers a, a5 and a3. The bold letters are the point mutations. The POT1-binding sites are denoted by boxes [29].
(b,c) Equilibrium binding curves for primers a3 (**b**) and a5 (**c**) binding to POT1 and the POT1-TPP1 complex. The solid and dashed lines represent theoretical binding curves fit to the data for POT1 and POT1-TPP1, respectively. The calculated equilibrium dissociation constant (*K*d) values are indicated. The binding curves and *K*d values for primers a, b and hT10 (TTAGGGTTAG) are shown in Fig. 2.3.

Figure 2.3 TPP1, POT1, and ssDNA form a ternary complex with enhanced stability relative to POT1-ssDNA.

(a, b, and c) Equilibrium binding curves for primers a ($(TTAGGG)_3$) (**a**), b ($(GGTTAG)_3$) (**b**) and hT10 (TTAGGGTTAG) (**c**) binding to POT1 and the POT1-TPP1 complex. The red and blue lines represent theoretical binding curves fit to the data for POT1 and POT1-TPP1, respectively. The calculated equilibrium dissociation constant (K_d) values are indicated.

(d,e) Kinetic stability of the POT1-TTP1-ssDNA and POT1-ssDNA complexes for primers a3 ($(TTTGGG(TTAGGG)_2$) (**d**) and a5 (TTAGGGTTAGCGTTAGGG) (**e**). Reactions containing 500 nM protein and 50 nM labeled ssDNA were incubated at room temperature for 30 min prior to challenge with 250-fold excess unlabeled ssDNA. Zero time points were taken prior to competitor addition. Bound and free DNA were separated by EMSA and quantified by PhosphorImager analysis. Data could not be adequately fit with a single exponential, so a double exponential fit was performed.


a
Binding curve (primer a)
Fraction bound (%)
Protein (nM)
POT1 K_D = 26 nM
POT1-TPP1 K_D = 4.3 nM
b
Binding curve (Primer b)
Fraction bound (%)
Protein (nM)
POT1 K_D = 6.7 nM
POT1-TPP1 K_D = 2.2 nM
c
Binding curve (Primer hT10)
Fraction bound (%)
Protein (nM)
POT1 K_D = 13 nM
POT1-TPP1 K_D = 3.3 nM
d
Cold chase (primer a3)
Fraction bound (%)
Time (min)
POT1-TPP1
POT1
e
Cold chase (primer a5)
Fraction bound (%)
Time (min)
POT1-TPP1
POT1

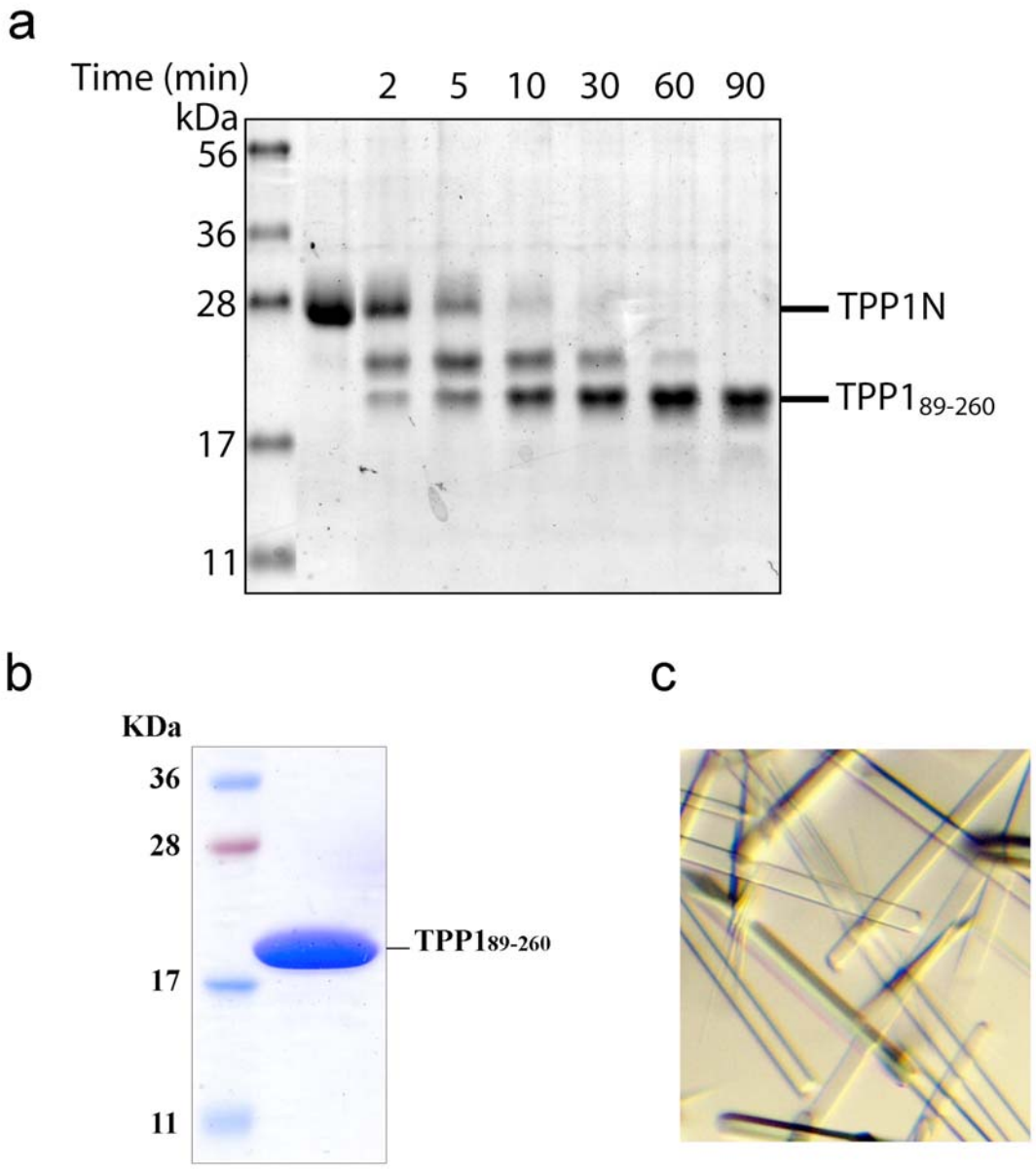


Figure 2.4 Determination of TPP1-OB from TPP1N
(a) SDS–PAGE time course of limited protease (Glu-C) cleavage of TPP1N. Lanes in minutes of time of the reaction as labeled.
(b) SDS-PAGE of purified $TPP1_{89-260}$ protein
(c) Crystals of $TPP1_{89-260}$

Table 2.1, Data collection, phasing and refinement statistics of TPP1-OB

	Native	Hg peak
Data collection		
Space group	$I4_122$	$I4_122$
Cell dimensions		
a, *b*, *c* (Å)	117.246, 117.246, 171.585	119.199, 119.199, 173.562
α, β, γ (°)	90, 90, 90	90, 90, 90
Wavelength (Å)	0.97950	1.00695
Resolution (Å) (high res. shell)	50-2.7 (2.8-2.7)	50-3.2 (3.31-3.20)
R_{merge} (%) (high res. shell)	11.6 (38.5)	14.7 (40.0)
I/σ (high res. shell)	25.2 (2.7)	16.5 (3.6)
Completeness (%) (high res. shell)	97.2 (81.7)	99.6 (96.4)
Redundancy (high res. shell)	9.8 (4.3)	8.3 (6.2)
Refinement		
Resolution (Å)	50-2.7	
No. reflections	16336	
R_{work}/R_{free} (%)	21.7/25.8	
No. atoms		
Protein	2344	
Water	83	
B-factors (Å^2)		
Protein	56.2	
Water	43.6	
R.m.s deviations		
Bond lengths (Å)	0.007694	
Bond angles (°)	1.69292	

Figure 2.5 The crystal structure of TPP1-OB indicates that TPP1 is the homologue of *O. nova* TEBPβ.
(a) Ribbon diagram of TPP1-OB with b-strands coloured in blue, a-helices green, and loops orange. The secondary structure elements are labelled.
(b) Superposition of TPP1-OB on the crystal structure of the OB fold of TEBPβ [26]. TPP1 is in red and TEBPβ in blue.
(c) Superposition of TPP1-OB on the crystal structure of the first OB fold of TEBPα [26]. TPP1 is in red and TEBPα in cyan. Figures were generated by using the program Pymol (http://pymol.sourceforge.net).

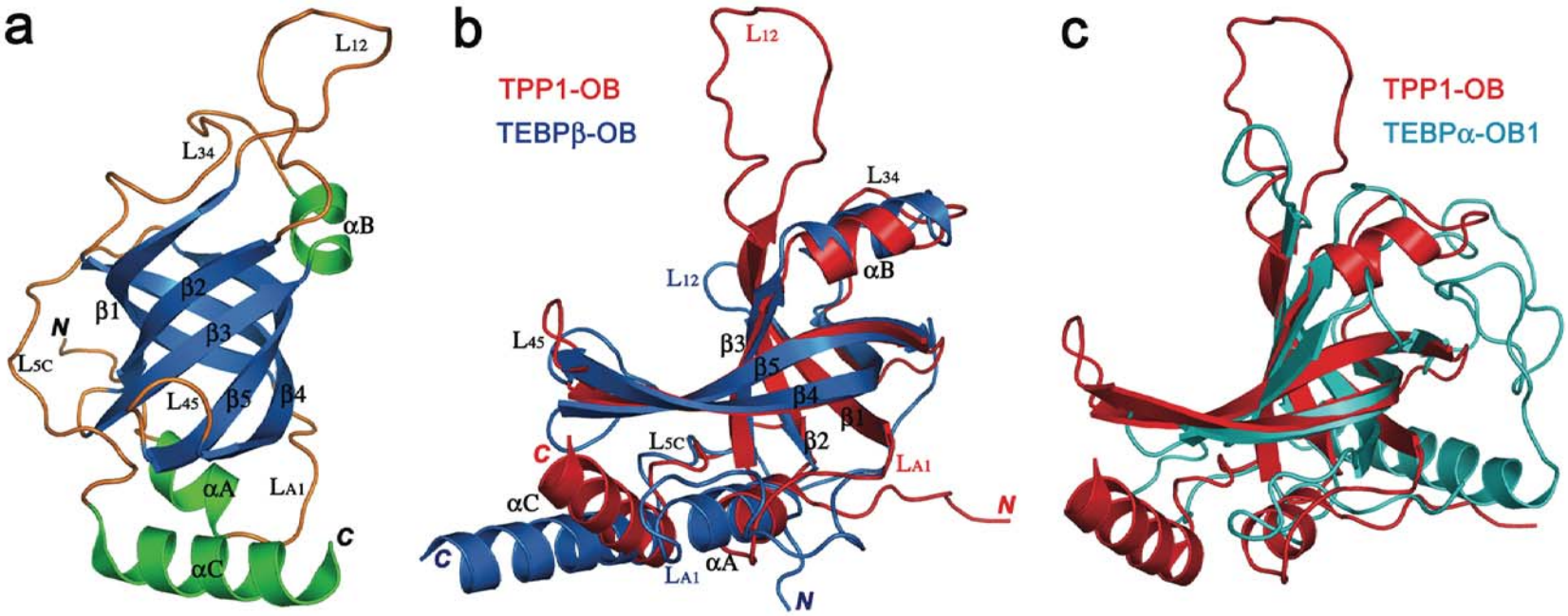

a
L12
L34
αB
β2
β1
N
β3
L5C
L45
β5
β4
αA
LA1
αC
C
b
TPP1-OB
TEBPβ-OB
L12
L34
αB
L12
L45
β3
β5
β4
β1
L5C
β2
C
LA1
αC
N
αA
C
LA1
N
c
TPP1-OB
TEBPα-OB1

Figure 2.6 Structure-based sequence alignment of the OB-folds of human TPP1 and its homologues.

Secondary structure assignments from the TPP1 crystal structure are shown as colored cylinders (α helices) and arrows (β strands) above the aligned sequences. Red dots indicate the conserved residues in the OB folds of TPP1 and TEBPβ. The dotted lines show the disordered regions.

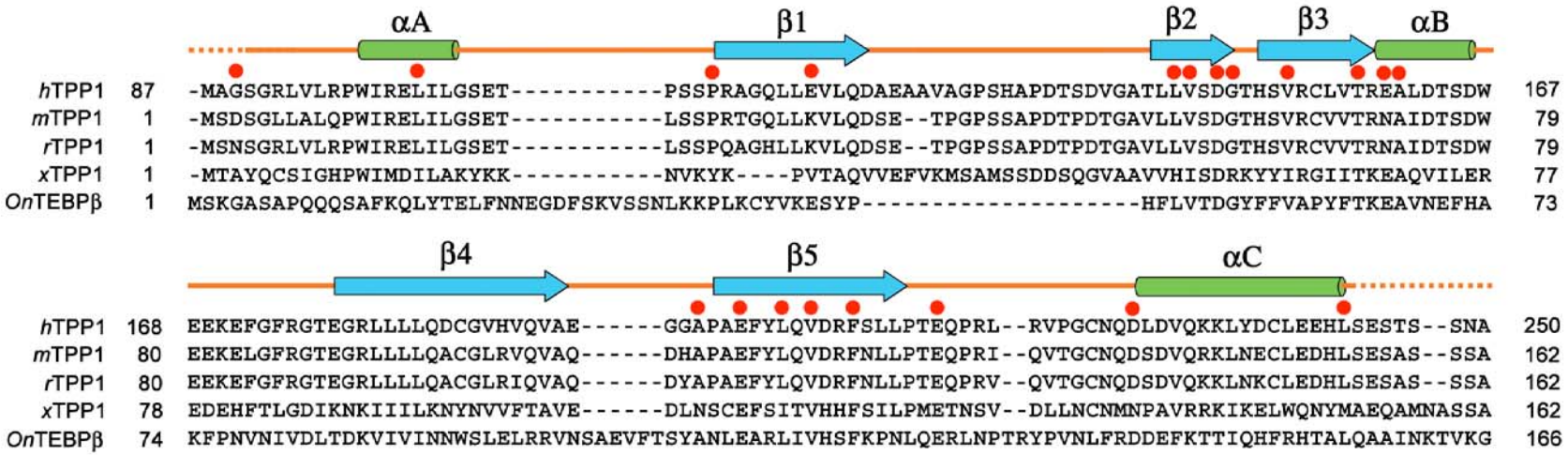
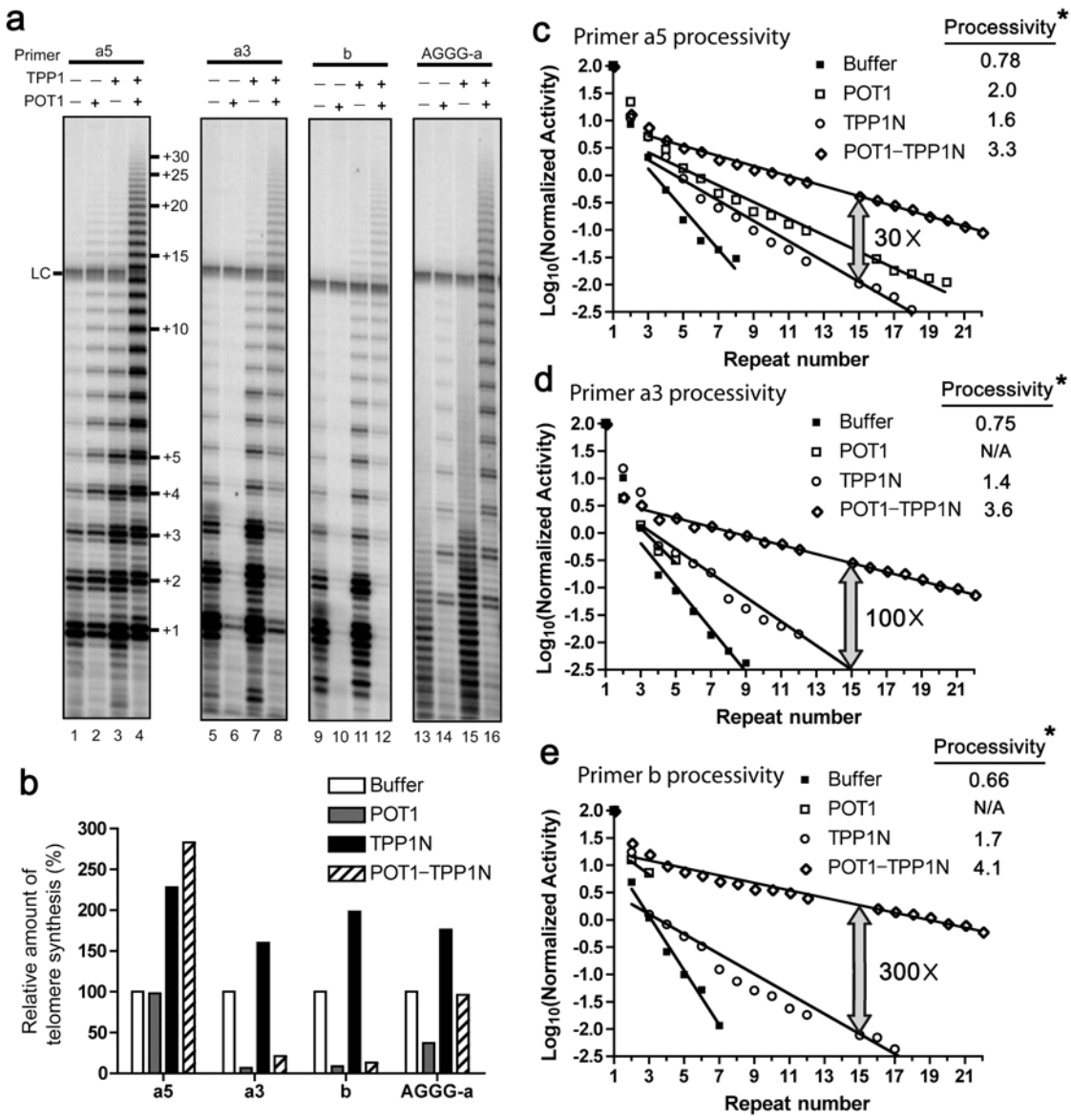

αA
β1
β2
β3
αB
hTPP1 87 -MAGSGRLVLRPWIRELILGSET----------PSSPRAGQLLEVLQDAEAAVAGPSHAPDTSDVGATLLVSDGTHSVRCLVTREALDTSDW 167
mTPP1 1 -MSDSGLLALQPWIRELILGSET----------LSSPRTGQLLKVLQDSE--TPGPSSAPDTPDTGAVLLVSDGTHSVRCVVTRNAIDTSDW 79
rTPP1 1 -MSNSGRLVLRPWIRELILGSET----------LSSPQAGHLLKVLQDSE--TPGPSSAPDTPDTGAVLLVSDGTHSVRCVVTRNAIDTSDW 79
xTPP1 1 -MTAYQCSIGHPWIMDILAKYKK----------NVKYK----PVTAQVVEFVKMSAMSSDDSQGVAAVVHISDRKYYIRGIITKEAQVILER 77
OnTEBPβ 1 MSKGASAPQQQSAFKQLYTELFNNEGDFSKVSSNLKKPLKCYVKESYP--------------------HFLVTDGYFFVAPYFTKEAVNEFHA 73
β4
β5
αC
hTPP1 168 EEKEFGFRGTEGRLLLLQDCGVHVQVAE------GGAPAEFYLQVDRFSLLPTEQPRL--RVPGCNQDLDVQKKLYDCLEEHLSESTS--SNA 250
mTPP1 80 EEKELGFRGTEGRLLLLQACGLRVQVAQ------DHAPAEFYLQVDRFNLLPTEQPRI--QVTGCNQDSDVQRKLNECLEDHLSESAS--SSA 162
rTPP1 80 EEKEFGFRGTEGRLLLLQACGLRIQVAQ------DYAPAEFYLQVDRFNLLPTEQPRV--QVTGCNQDSDVQKKLNKCLEDHLSESAS--SSA 162
xTPP1 78 EDEHFTLGDIKNKIIILKNYNVVFTAVE------DLNSCEFSITVHHFSILPMETNSV--DLLNCNMNPAVRRKIKELWQNYMAEQAMNASSA 162
OnTEBPβ 74 KFPNVNIVDLTDKVIVINNWSLELRRVNSAEVFTSYANLEARLIVHSFKPNLQERLNPTRYPVNLFRDDEFKTTIQHFRHTALQAAINKTVKG 166

Figure 2.7 The POT1–TPP1 complex functions as a telomerase processivity factor.
(a) Direct telomerase activity assays [36] were performed with 100 nM primer a5 (lanes 1–4), a3 (lanes 5–8), b (lanes 9–12), or AGGG-a (lanes 13–16) in the presence of a saturating concentration of POT1, TPP1, or POT1–TPP1. Reaction products were then analyzed by gel electrophoresis (LC, loading control). Three independent sets of experiments gave equivalent results.
(b) Quantification of total DNA synthesis relative to synthesis in the presence of protein buffer alone.
(c–e) Activity in each repeat shown in **(a)** was measured, corrected for the number of radiolabelled nucleotides incorporated, and then plotted.
*Processivity= $R_{1/2}$= -ln2/(2.303k), where k is the slope and $R_{1/2}$ is the number of repeats synthesized before half of the chains have dissociated, analogous to $t_{1/2}$ radioactive decay.


a
Primer
a5
a3
b
AGGG-a
TPP1
POT1
LC
+30
+25
+20
+15
+10
+5
+4
+3
+2
+1
1 2 3 4
5 6 7 8
9 10 11 12
13 14 15 16
b
Buffer
POT1
TPP1N
POT1–TPP1N
Relative amount of telomere synthesis (%)
a5
a3
b
AGGG-a
c
Primer a5 processivity
Processivity*
Buffer 0.78
POT1 2.0
TPP1N 1.6
POT1–TPP1N 3.3
30X
Log10(Normalized Activity)
Repeat number
d
Primer a3 processivity
Processivity*
Buffer 0.75
POT1 N/A
TPP1N 1.4
POT1–TPP1N 3.6
100X
e
Primer b processivity
Processivity*
Buffer 0.66
POT1 N/A
TPP1N 1.7
POT1–TPP1N 4.1
300X

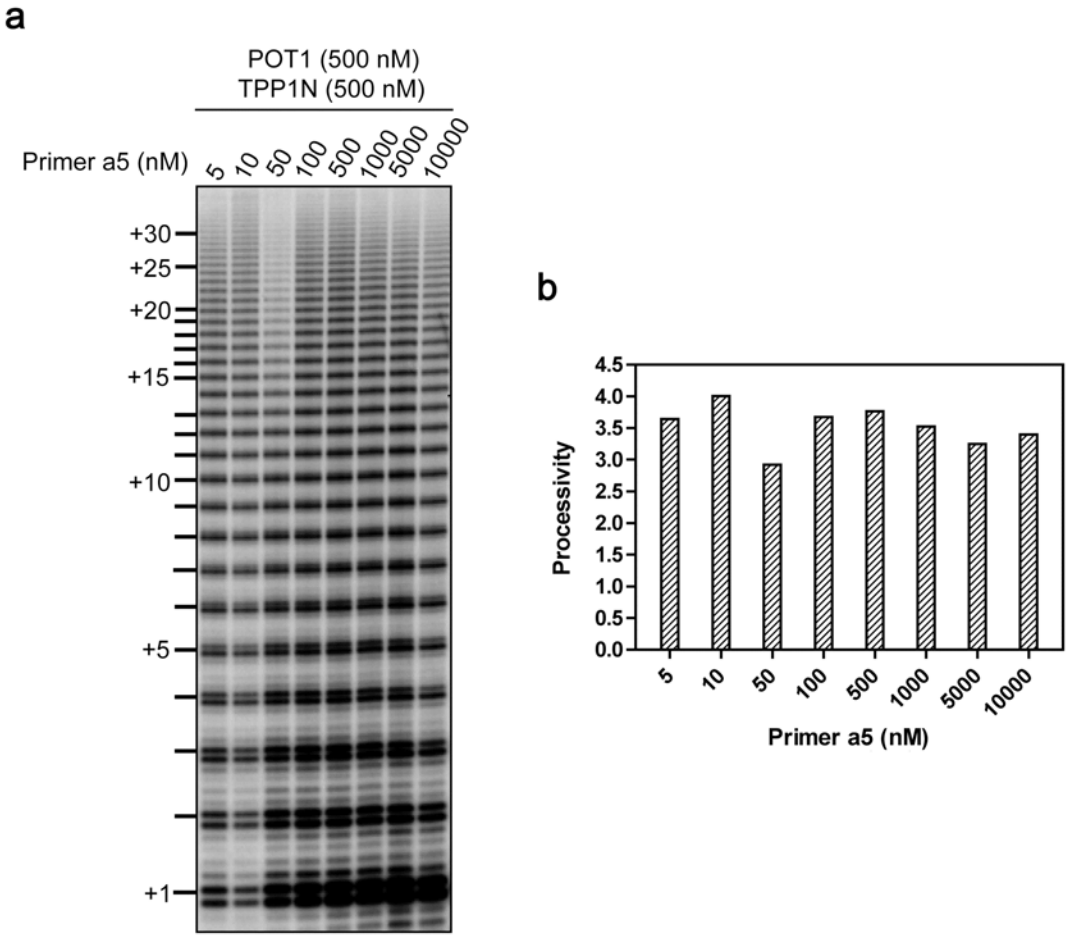


Figure 2.8 Effect of primer concentration on the processivity of telomerase.
(a) Direct telomerase activity assays were performed under standard conditions with primer a5 at concentrations of 5, 20, 50, 100, 500, 1,000, 5,000 and 10,000 nM in the presence of POT1 (500 nM) and TPP1 (500 nM). Reaction products were then analyzed by gel electrophoresis.
(b) A bar-representation of the telomerase processivity values**.**
*Processivity= $R_{1/2}$= -ln2/(2.303k), where k is the slope and $R_{1/2}$ is the number of repeats synthesized before half of the chains have dissociated, analogous to $t_{1/2}$ radioactive decay.

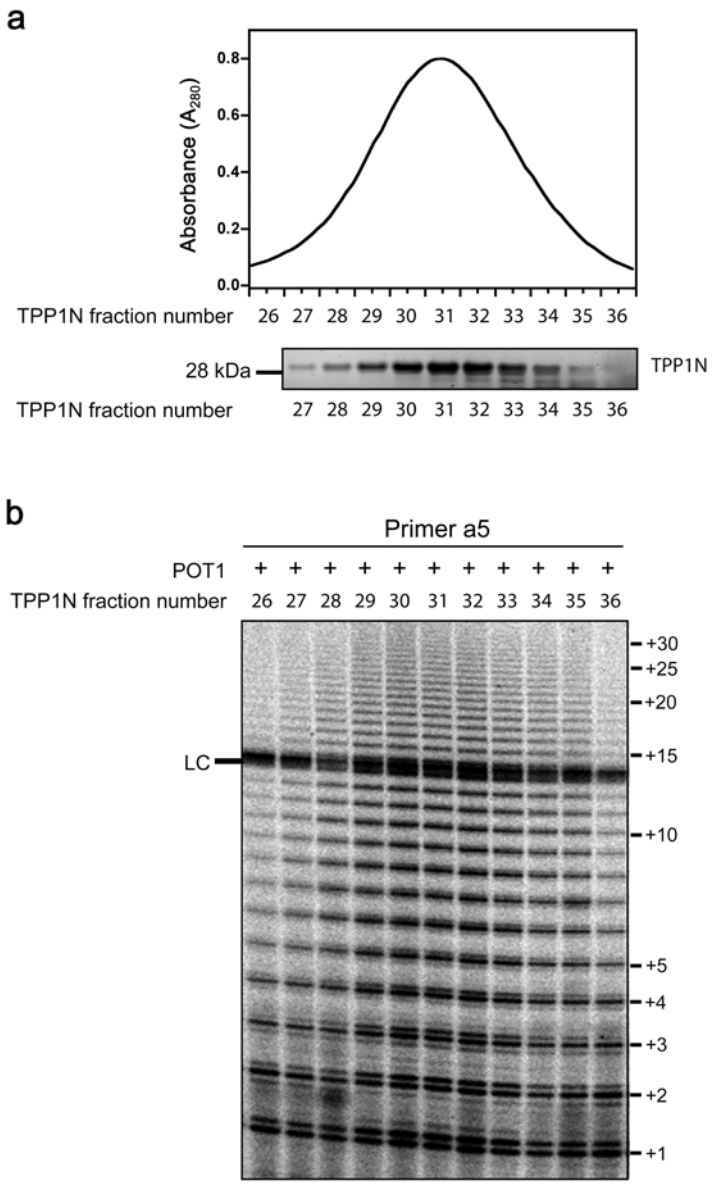
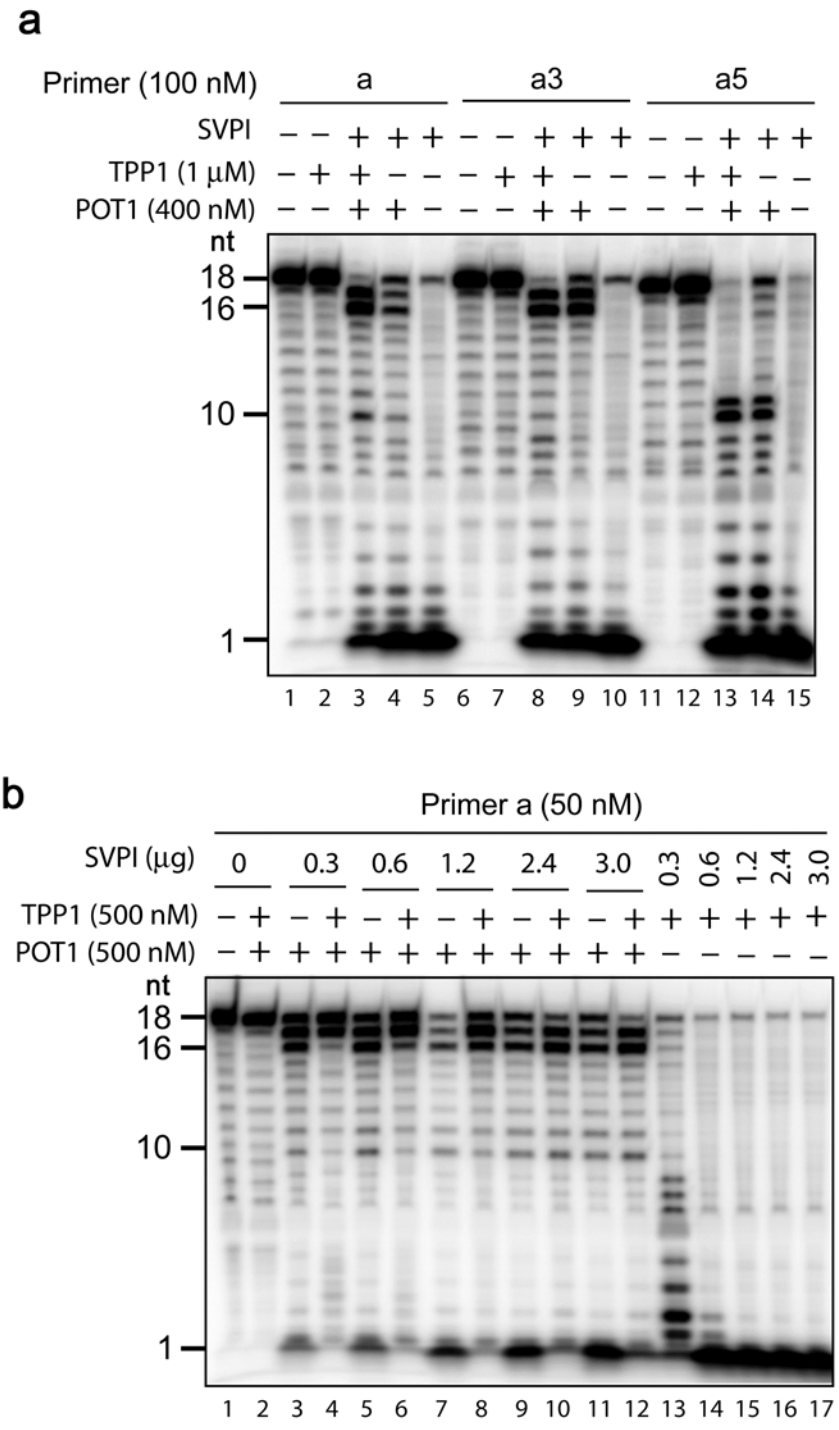


Figure 2.9 The enhanced processivity of telomerase is TPP1 specific.
(a) Upper panel: Gel filtration chromatography profile of TPP1N. Lower panel: SDS-PAGE gel of TPP1 fractions corresponding to the peak of TPP1N in gel filtration profile. **(b)** Direct telomerase activity assays were performed using 100 nM primer a5 and an equimolar amount of purified POT1 protein with the central TPP1N peak fractions in panel a.

Figure 2.10 TPP1 does not substantially relocate POT1 to an internal site on the DNA.
(a) SVPI digestion of primers a, a3, and a5 in the presence of protein buffer (-), TPP1, POT1, or POT1-TPP1 was carried out under the same conditions as that of the direct telomerase activity assay.
(b) SVPI digestion results of primer a in the presence of TPP1, POT1, or POT1-TPP1 with increasing amount of SVPI. TPP1 binds to POT1-ssDNA and enhances the POT1-ssDNA interaction (compare the amounts of the 1 nt degradation products in lanes 3, 5, 7, 9, 11 and 4, 6, 8, 10, 12). TPP1 by itself didn't protect DNA from SVPI digestion (lanes 13-17).


a
Primer (100 nM)
a
a3
a5
SVPI
TPP1 (1 μM)
POT1 (400 nM)
nt
18
16
10
1
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
b
Primer a (50 nM)
SVPI (μg)
0
0.3
0.6
1.2
2.4
3.0
0.3
0.6
1.2
2.4
3.0
TPP1 (500 nM)
POT1 (500 nM)
nt
18
16
10
1
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17

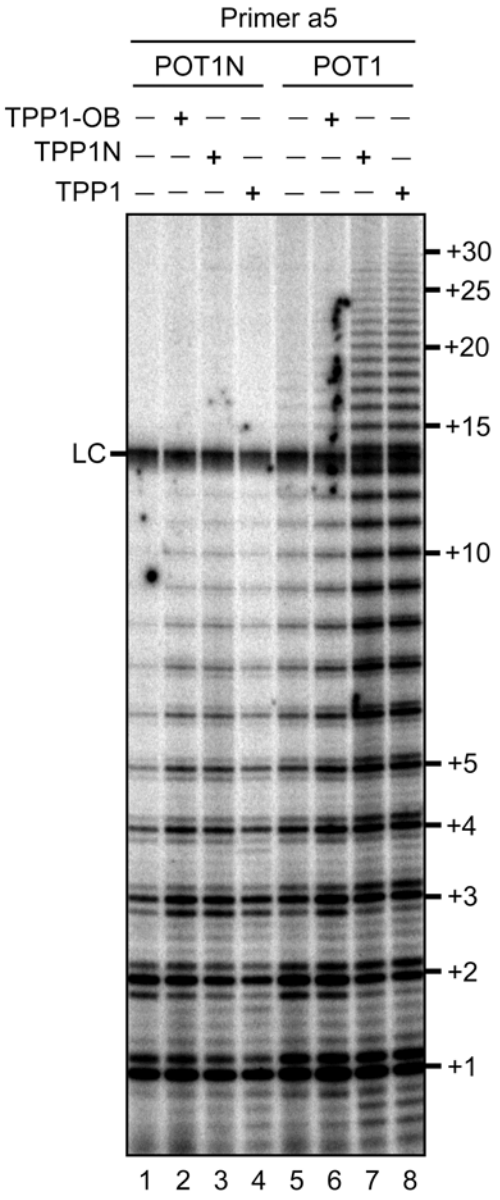


Figure 2.11 POT1-TPP1 interaction is necessary for the enhanced processivity and activity of telomerase.
Direct telomerase activity assays were performed using 100 nM primer a5 with protein buffer (-), TPP1-OB, TPP1N, or TPP1 in the presence of POT1N (lanes 1-4) or POT1 (lanes 5-8). All proteins were present at saturating concentrations.

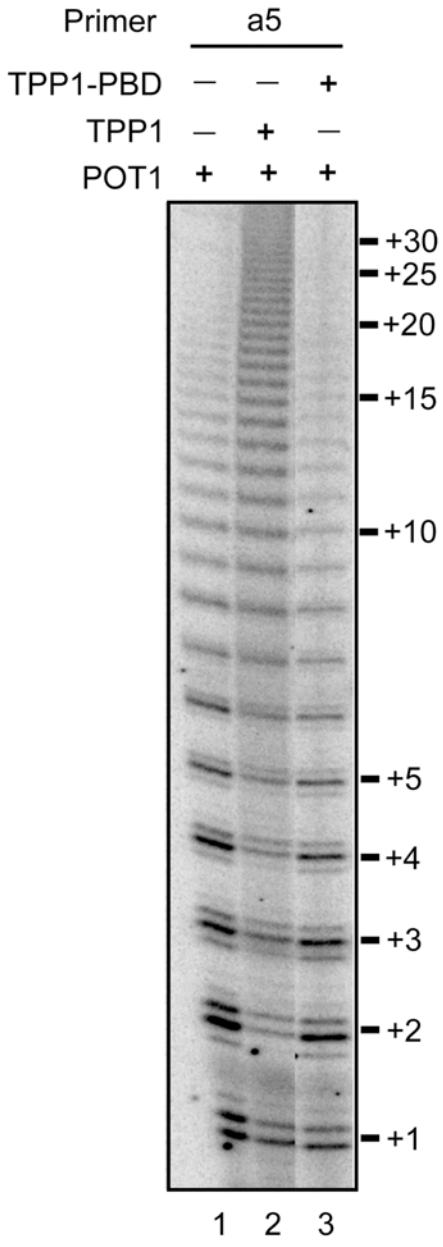
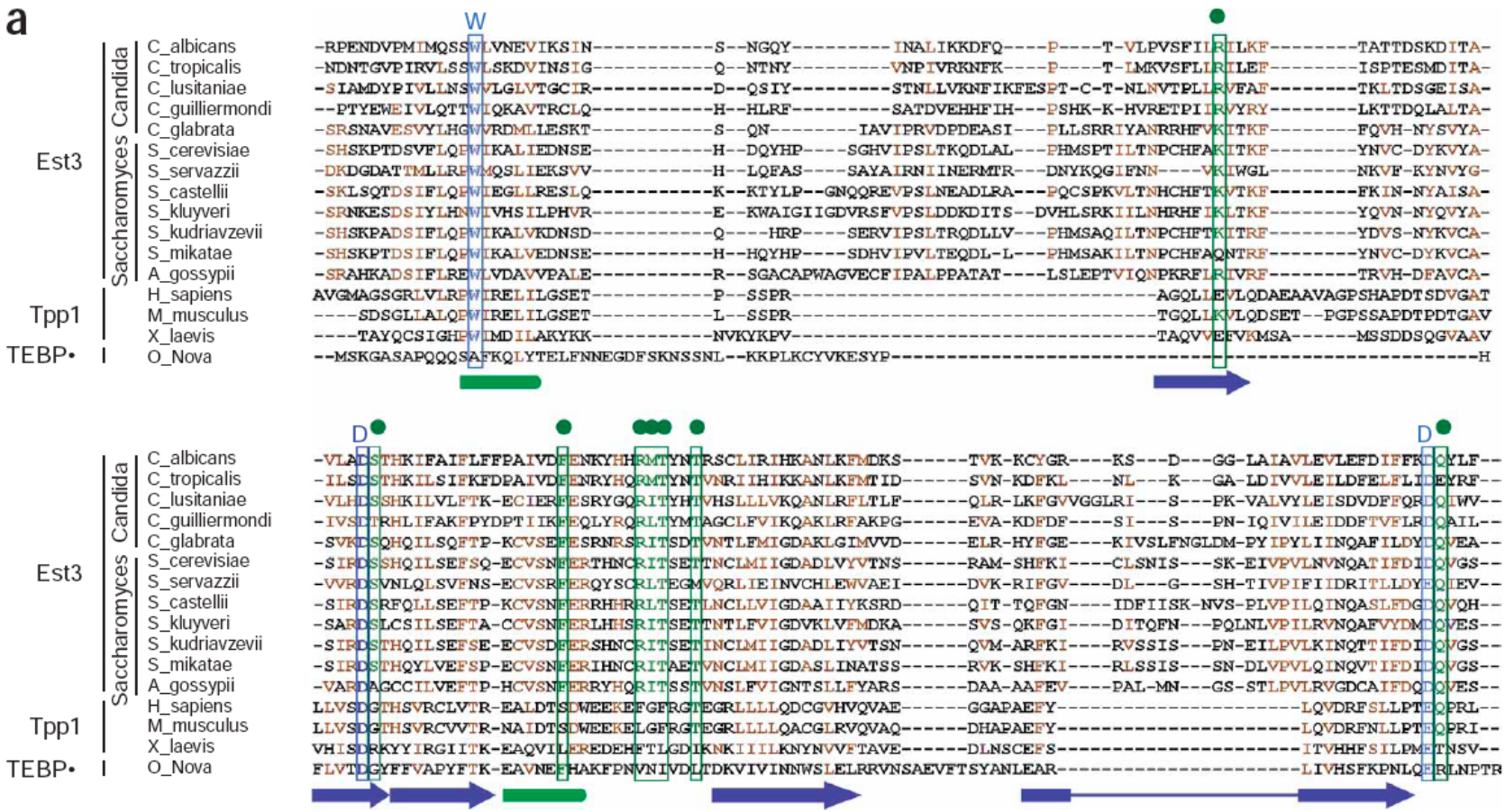


Figure 2.12 The POT1-binding domain of TPP1, TPP1-PBD, is insufficient to activate telomerase in the presence of POT1.
Direct telomerase activity assays were performed using 100 nM primer a5 with protein buffer (-), TPP1-PBD, or TPP1 in the presence of POT1.

Figure 2.13 Sequence and structural similarity between Est3 and Tpp1 homologs. **(a)** Sequence comparison between Est3, TPP1 and TEBPβ. The secondary- structural elements derived from the human TPP1 crystal structure are illustrated as follows: α-helices, green cylinders; β-strands, blue arrows. Blue and cyan letters highlight mutated residues that are conserved between Est3 and TPP1. Note that these residues are also conserved in the recently identified *S. pombe* homolog Tpz1 (ref. [64] and data not shown). Green dots highlight mutated residues that are conserved only in Est3.
(b) A hypothetical structure of Est3 based on the TPP1 crystal structure. Residues predicted to be important for telomerase interaction based on structure-function analysis are shown in van der Waals sphere representations in blue (Asp91) and cyan (Trp36 and Asp169).


a
W
D
D
Est3
Tpp1
TEBP•
Saccharomyces
Candida
C_albicans
C_tropicalis
C_lusitaniae
C_guilliermondi
C_glabrata
S_cerevisiae
S_servazzii
S_castellii
S_kluyveri
S_kudriavzevii
S_mikatae
A_gossypii
H_sapiens
M_musculus
X_laevis
O_Nova

b

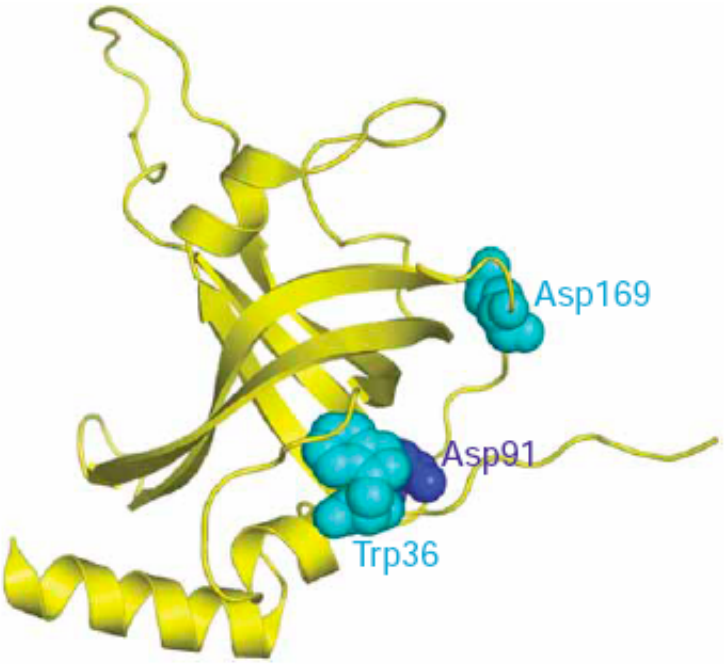
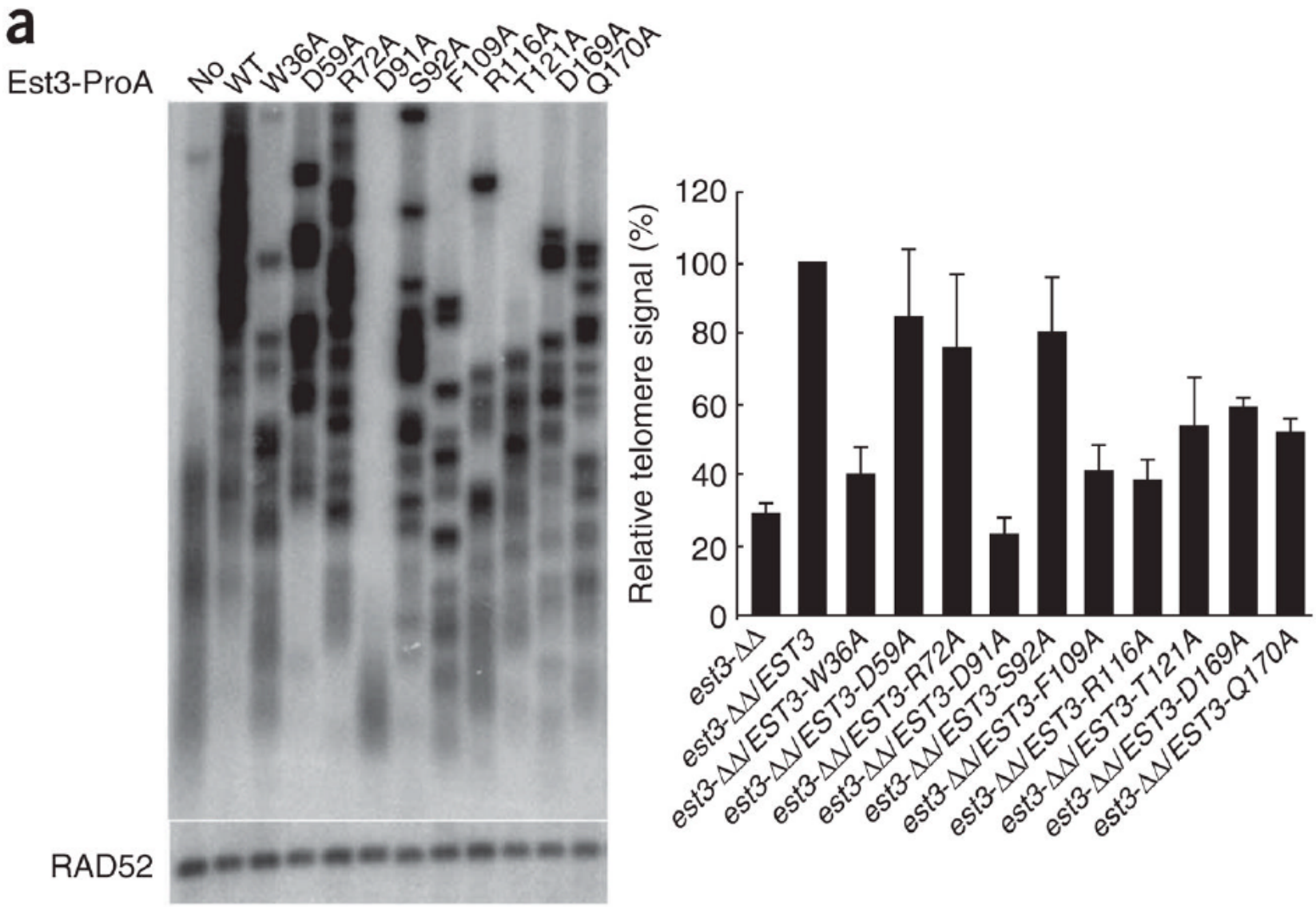
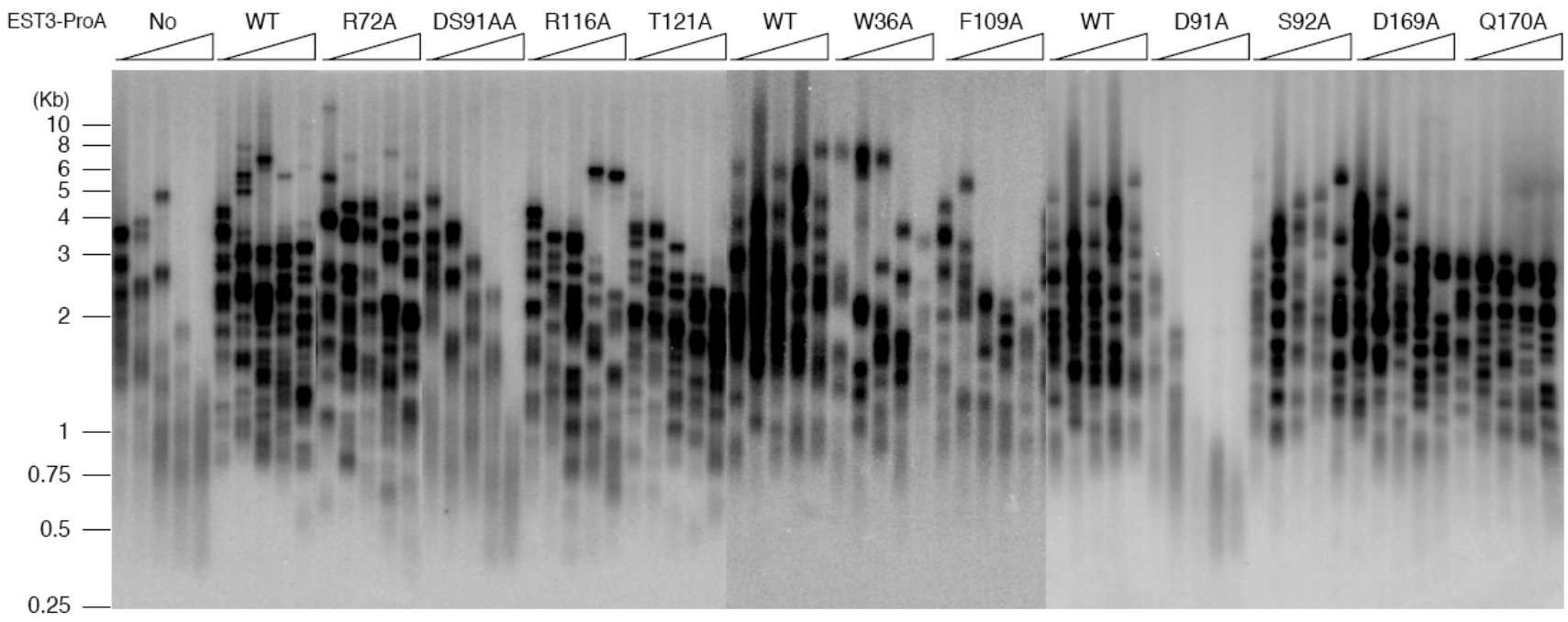


Figure 2.14 The effects of *C. albicans* Est3 mutations on telomere maintenance and telomerase association. (Done by Dr. Lue's group)
(a) Left, *est3*Δ/*est3*Δ strains containing the wild-type and mutant alleles of protein A (ProA)- tagged *EST3* were passaged on plates by successively streaking for single colonies. Chromosomal DNAs were prepared from either streaks 16 or 20 of each indicated strain and subjected to telomere Southern blotting (above). The blot was then stripped and reprobed using a labeled *RAD52* fragment (below). Right, the telomere-hybridization signals and the *RAD52* signals obtained from the Southern blots shown and from three other independent blots were quantified using a PhosphorImager. The relative ratios of the telomere to *RAD52* signals were calculated and plotted for the wild-type and all mutant samples. Shown are means ± s.d. from four independent experiments.
(b) Effects of *Candida* EST3 mutations on telomere lengths. Telomeres from successive streaks of *est3*Δ/ *est3*Δ strains with different mutant alleles of protein A-tagged *EST3* were analyzed by Southern blotting. The chromosomal DNAs were derived from streaks 4, 8, 12, 16 and 20 of each indicated strain. One streak corresponds to ~ 25 generations of growth. The results are representative of at least two independent clones.
(c) Est3 and associated Ter1 in wild-type and mutant strains were analyzed by IgG-Sepharose pull-down followed by western analysis (right) and RT-PCR (20 cycles) (left), respectively.
(d) Extracts from strains containing ProA-tagged wild- type *EST3* (WT), the W36A mutant and the D169A mutant were subjected to IgG-Sepharose pull-down and RT-PCR to detect Est3-associated Ter1. To facilitate quantitative comparison, the WT sample was subjected to three-fold serial dilution before RT-PCR. Thermocycling was performed for 20 or 23 cycles to ensure the linearity of the signals.
(e) The Est3 protein and associated Ter1 in the wild-type and D91A mutant strains were analyzed by IgG-Sepharose pull-down followed by western blotting and RT-PCR (below and above, respectively). The WT sample was subjected to five-fold serial dilution before the pull-down.


a
Est3-ProA
No
WT
W36A
D59A
R72A
D91A
S92A
F109A
R116A
T121A
D169A
Q170A
RAD52
Relative telomere signal (%)
120
100
80
60
40
20
0
est3-ΔΔ
est3-ΔΔ/EST3
est3-ΔΔ/EST3-W36A
est3-ΔΔ/EST3-D59A
est3-ΔΔ/EST3-R72A
est3-ΔΔ/EST3-D91A
est3-ΔΔ/EST3-S92A
est3-ΔΔ/EST3-F109A
est3-ΔΔ/EST3-R116A
est3-ΔΔ/EST3-T121A
est3-ΔΔ/EST3-D169A
est3-ΔΔ/EST3-Q170A

b


EST3-ProA
No
WT
R72A
DS91AA
R116A
T121A
WT
W36A
F109A
WT
D91A
S92A
D169A
Q170A
(Kb)
10
8
6
5
4
3
2
1
0.75
0.5
0.25

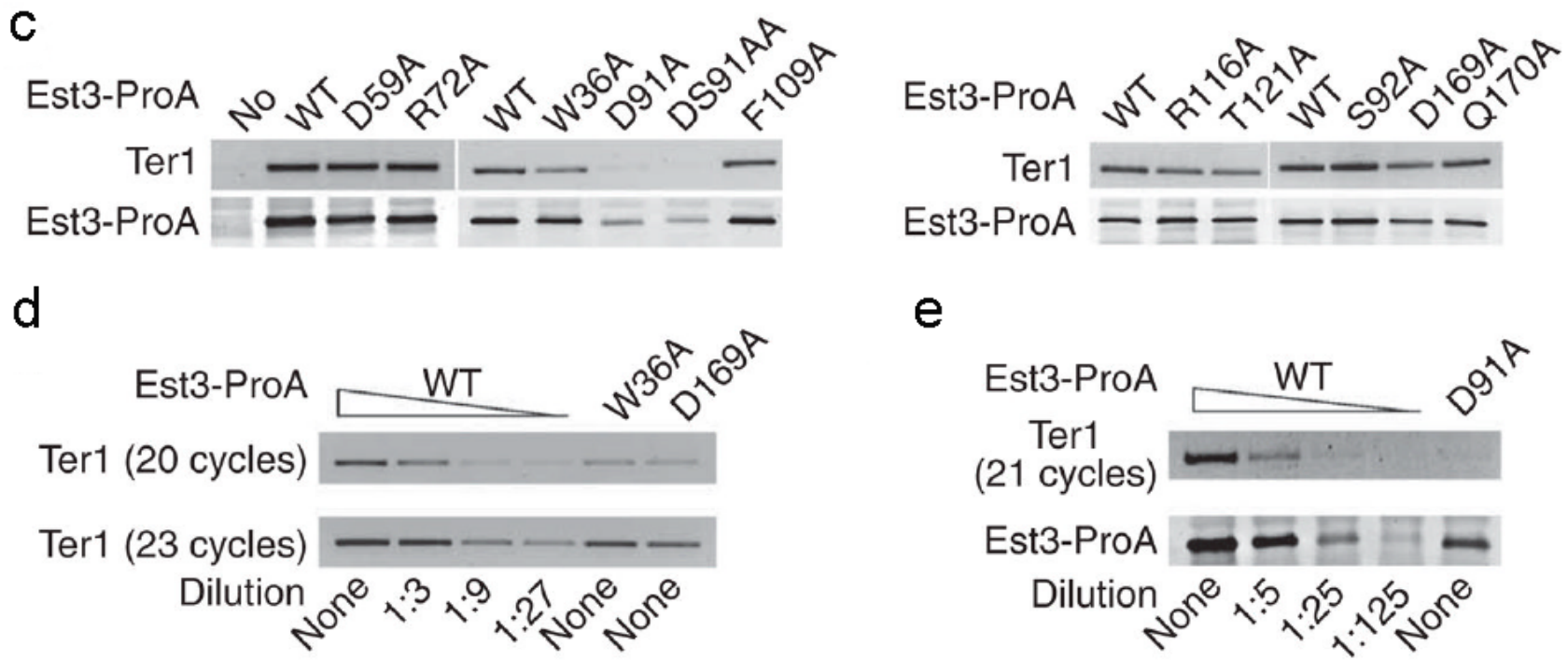

c
Est3-ProA
No
WT
D59A
R72A
WT
W36A
D91A
DS91AA
F109A
Ter1
Est3-ProA
Est3-ProA
WT
R116A
T121A
WT
S92A
D169A
Q170A
Ter1
Est3-ProA
d
Est3-ProA
WT
W36A
D169A
Ter1 (20 cycles)
Ter1 (23 cycles)
Dilution
None
1:3
1:9
1:27
None
None
e
Est3-ProA
WT
D91A
Ter1
(21 cycles)
Est3-ProA
Dilution
None
1:5
1:25
1:125
None

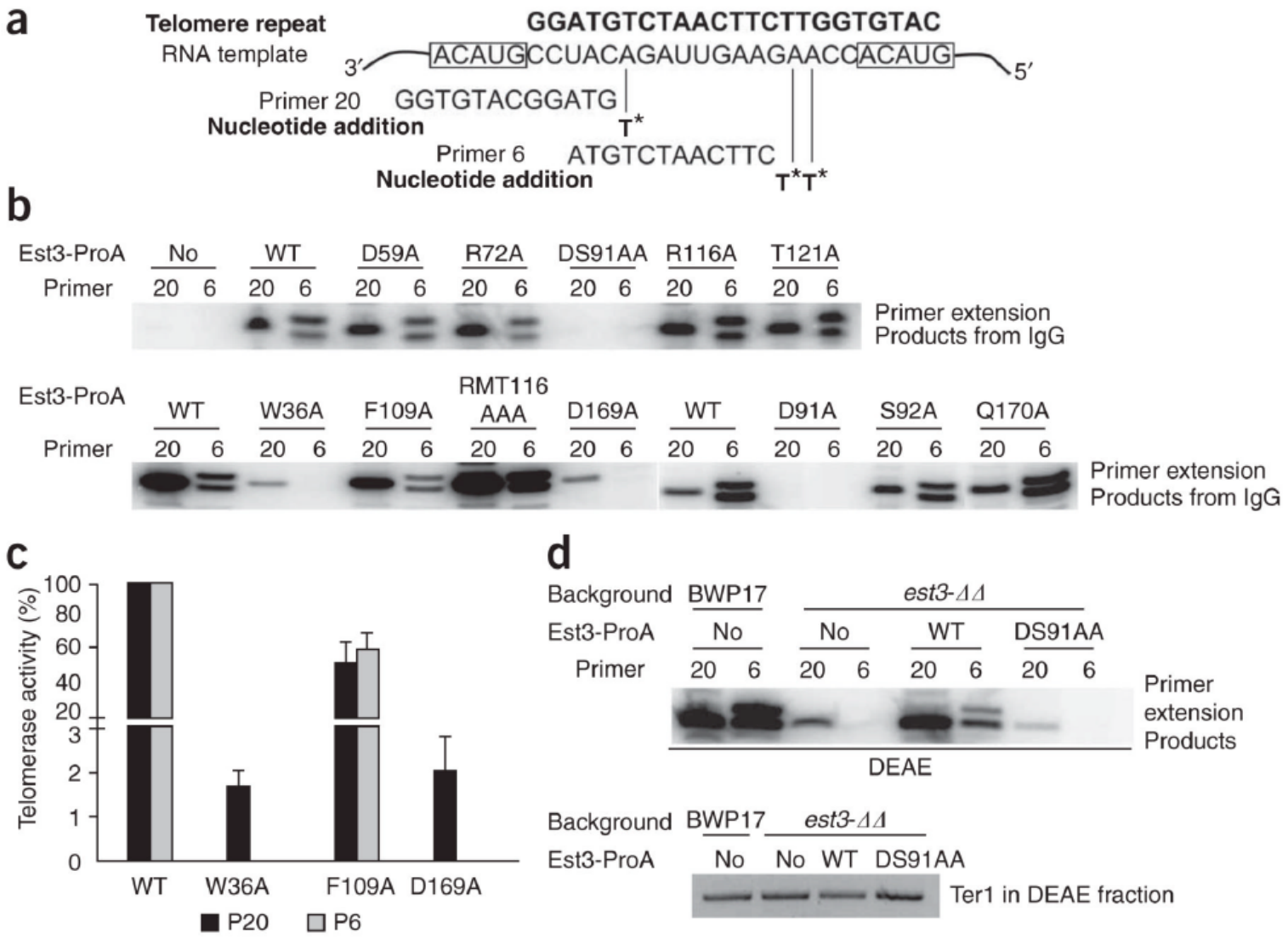
Figure 2.15 The effects of *C. albicans* Est3 mutations on telomerase primer extension activity *in vitro*. (Done by Dr. Lue's group)

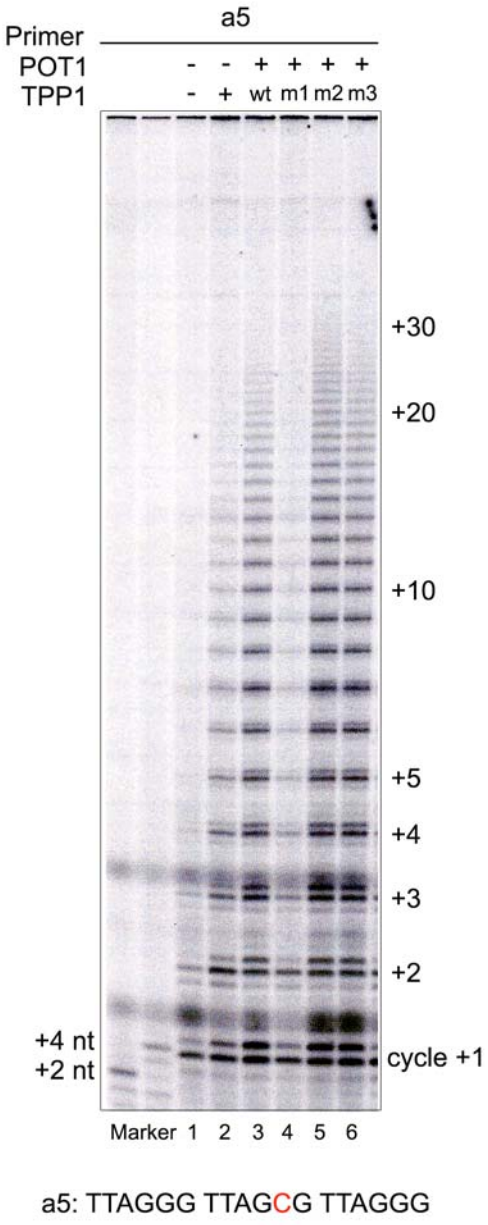
(a) All primer extension assays were performed using two different 12-nt primers, P6 and P20, which correspond to different regions of the *C. albicans* telomere repeat and support the synthesis of primer+1 or primer+2 products in the presence of labeled dTTP (T*).

(b) Telomerase isolated from the indicated strains by IgG-Sepharose pull-down were assayed for primer extension activity using primers P6 and P20 in the presence of labeled dTTP.

(c) The activities for the W36A, F109A and D169A mutants relative to the wild-type enzyme on the P6 and P20 primer were quantified and plotted; shown are means ± s.d. from three independent experiments.

(d) Telomerase isolated by DEAE chromatography from the indicated strains was tested for enzyme activity using primers P6 and P20 in the presence of labeled dTTP (above). The levels of Ter1 in the fractions were measured by RT-PCR (below).


a
Telomere repeat
RNA template
GGATGTCTAACTTCTTGGTGTAC
3′
ACAUGCCUACAGAUUGAAGAACCACAUG
5′
Primer 20
GGTGTACGGATG
Nucleotide addition
T*
Primer 6
ATGTCTAACTTC
Nucleotide addition
T*T*
b
Est3-ProA
No
WT
D59A
R72A
DS91AA
R116A
T121A
Primer
20 6 20 6 20 6 20 6 20 6 20 6 20 6
Primer extension Products from IgG
Est3-ProA
WT
W36A
F109A
RMT116 AAA
D169A
WT
D91A
S92A
Q170A
Primer
20 6 20 6 20 6 20 6 20 6 20 6 20 6 20 6 20 6
Primer extension Products from IgG
c
Telomerase activity (%)
100
80
60
40
20
3
2
1
0
WT
W36A
F109A
D169A
P20
P6
d
Background
BWP17
est3-ΔΔ
Est3-ProA
No
No
WT
DS91AA
Primer
20 6 20 6 20 6 20 6
Primer extension Products
DEAE
Background
BWP17
est3-ΔΔ
Est3-ProA
No
No
WT
DS91AA
Ter1 in DEAE fraction



2.16 Effects of TPP1 mutations on telomerase activity

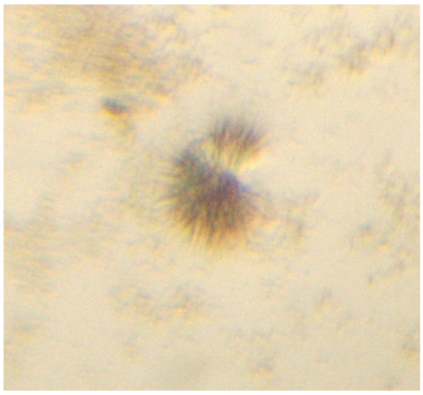
Three mutant and the wild type TPP1N proteins are used in the *in vitro* telomerase activity assay. When m1: W98A (lane 4) was introduced, it greatly abolished POT1-TPP1 complex's processive effect on telomerase. However, the other two mutants (m2: E215A and m3: E215A/Q216A, lane 5 and 6) didn't have such effect. The DNA primer was extended just like the wild type (lane 3) did.

Table 2.2 Phenotypes of the EST3 mutants

Class	Mutated residues	Telomere maintenance	Telomerase association	Telomerase activity
I	Asp91	Similar to deletion mutant	No	No
II	Trp36, Asp169	Moderate and progressive loss	Reduced (~5x)	Reduced (~50x)
III	Phe109	Moderate and progressive loss	Normal	Reduced (~2x)
	Arg116, Thr121	Moderate and progressive loss	Normal	Normal
	Gln170	Short and stable	Normal	Normal
IV	Asp59, Arg72, Ser92	Similar to wild type	Normal	Normal

Figure 2.17 Structural study (crystallography and NMR) of Est3
(a) *Sc*Est3 (2-181) crystallizes in the condition of HEPES pH 7.5 100 mM, PEG 4000 30%, NaCl 200 mM.
(b) The ^{1}H-^{15}N 2D spectrum analysis for the ^{15}N-labeled *Candida. tropicalis* Est3 (residue 4-196). There are about 170 ^{1}H-^{15}N isolated peaks nicely distributed on the spectrum after 3 hours of data collection. (Done by Dr. Erik Zuiderweg's group)

a



b

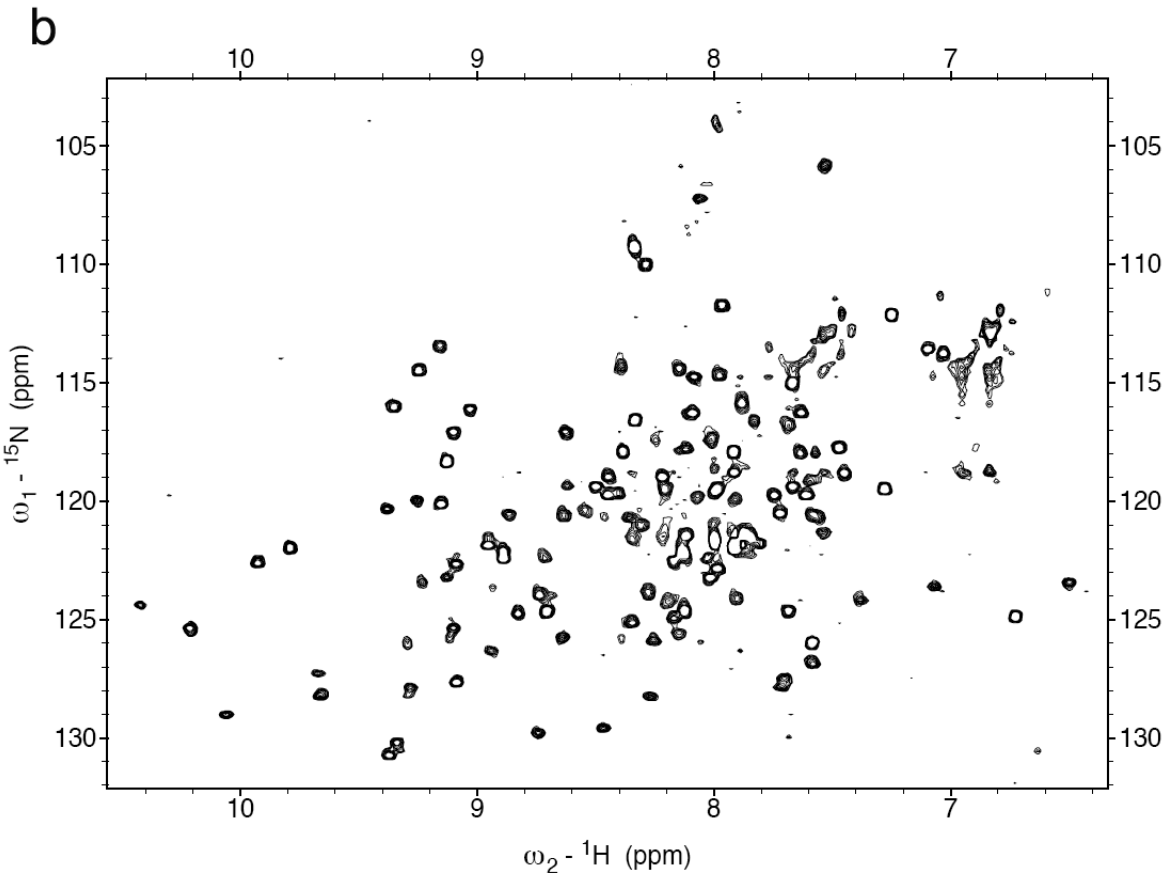
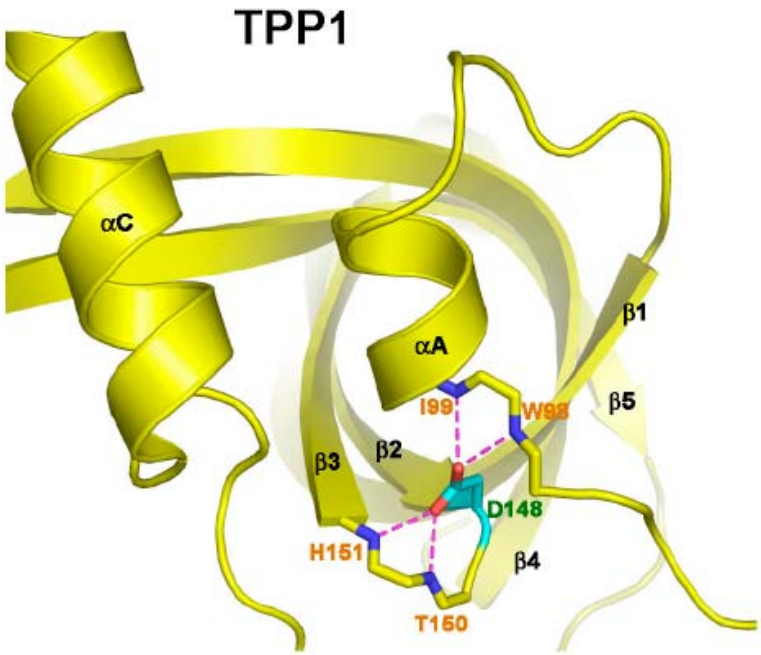


Figure 2.18 The conservation of an Asp residue in OB-fold proteins and its structural role.
(a) The backbone structure of TPP1 is shown in a ribbon representation. Residue D146, located at the end of β2, is shown as sticks that make four hydrogen bounding interactions with residues in the L2-3 loop (T150 and H151) and residues near αA (W98 and I99).
(b) A local alignment of sequences near the end of β2 from many OB-fold proteins (including Est3, TPP1, POT1, TEBPβ and RPA) illustrates the conservation of this aspartic acid residue.

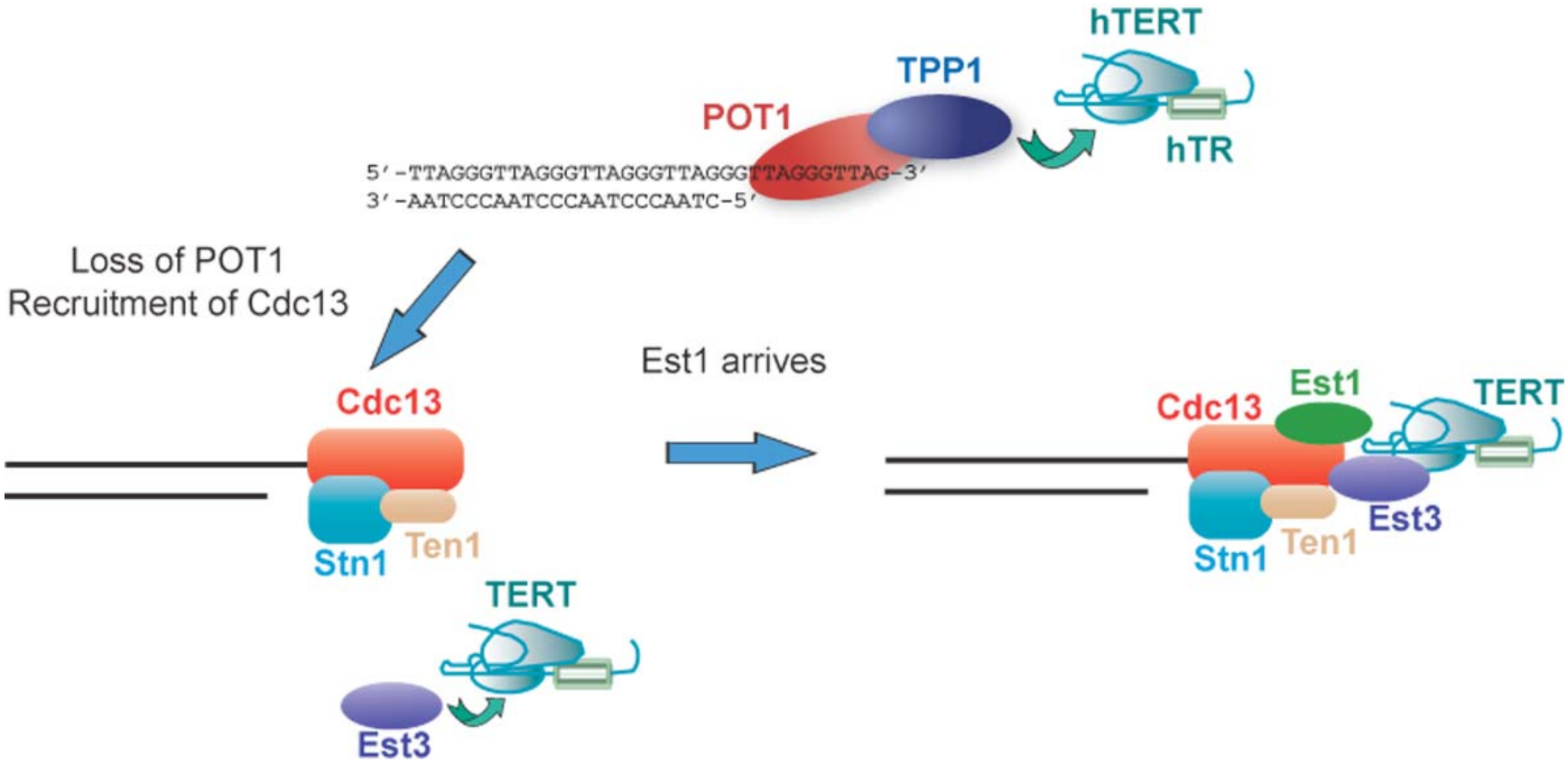
a



b

	β2
*Ca*Est3	ITAVIADSTH
hTPP1	ATLLVSDGTH
*On*TEBPβ	PHFLVTDGYF
hPOT1_OB1	SVVTIVDQTN
hPOT1_OB2	FLLKVWDGTR
*On*TEBPα_OB1	CSLKIVDPTL
*On*TEBPα_OB2	NELKLKDASG
*On*TEBPα_OB3	VQFLVKDAST
hRPA-A	FSLELVDESG
hRPA-B	IVYKIDDMTA
hRAP-C	KMFILSDGEG

Figure 2.19 A possible evolutionary scenario for yeast telomere binding proteins and telomere.
The precursor of TPP1/Est3 may be part of the telomeric protein complex and a stimulatory factor for telomerase. During budding yeast evolution, this precursor may be lost from the telomeres and its function in telomere protection and telomerase activation taken up by the CST complex and telomerase-bound Est3, respectively. See text fro more detailed explanations.


hTERT
TPP1
POT1
hTR
5′-TTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAG-3′
3′-AATCCCAATCCCAATCCCAATC-5′
Loss of POT1
Recruitment of Cdc13
Cdc13
Stn1
Ten1
TERT
Est3
Est1 arrives
Cdc13
Est1
TERT
Est3
Stn1
Ten1

Reference:

1. Baumann, P. and T.R. Cech, *Pot1, the putative telomere end-binding protein in fission yeast and humans.* Science, 2001. **292**(5519): p. 1171-5.
2. Baumann, P., E. Podell, and T.R. Cech, *Human Pot1 (protection of telomeres) protein: cytolocalization, gene structure, and alternative splicing.* Mol Cell Biol, 2002. **22**(22): p. 8079-87.
3. Tani, A. and M. Murata, *Alternative splicing of Pot1 (Protection of telomere)-like genes in Arabidopsis thaliana.* Genes Genet Syst, 2005. **80**(1): p. 41-8.
4. Wei, C. and C.M. Price, *Cell cycle localization, dimerization, and binding domain architecture of the telomere protein cPot1*. Mol Cell Biol, 2004. **24**(5): p. 2091-102.
5. Wu, L., et al., *Pot1 deficiency initiates DNA damage checkpoint activation and aberrant homologous recombination at telomeres.* Cell, 2006. **126**(1): p. 49-62.
6. Hockemeyer, D., et al., *Recent expansion of the telomeric complex in rodents: Two distinct POT1 proteins protect mouse telomeres.* Cell, 2006. **126**(1): p. 63-77.
7. Lei, M., et al., *DNA self-recognition in the structure of Pot1 bound to telomeric single-stranded DNA.* Nature, 2003. **426**(6963): p. 198-203.
8. Lei, M., E.R. Podell, and T.R. Cech, *Structure of human POT1 bound to telomeric single-stranded DNA provides a model for chromosome end-protection.* Nat Struct Mol Biol, 2004. **11**(12): p. 1223-9.
9. Houghtaling, B.R., et al., *A dynamic molecular link between the telomere length regulator TRF1 and the chromosome end protector TRF2*. Curr Biol, 2004. **14**(18): p. 1621-31.
10. Liu, D., et al., *PTOP interacts with POT1 and regulates its localization to telomeres.* Nat Cell Biol, 2004. **6**(7): p. 673-80.
11. Ye, J.Z., et al., *POT1-interacting protein PIP1: a telomere length regulator that recruits POT1 to the TIN2/TRF1 complex.* Genes Dev, 2004. **18**(14): p. 1649-54.
12. Blackburn, E.H., *Switching and signaling at the telomere.* Cell, 2001. **106**(6): p. 661-73.
13. Cech, T.R., *Beginning to understand the end of the chromosome.* Cell, 2004. **116**(2): p. 273-9.
14. Shay, J.W. and W.E. Wright, *Telomerase: a target for cancer therapeutics.* Cancer Cell, 2002. **2**(4): p. 257-65.
15. Chong, L., et al., *A human telomeric protein.* Science, 1995. **270**(5242): p. 1663-7.
16. Broccoli, D., et al., *Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2.* Nat Genet, 1997. **17**(2): p. 231-5.
17. Loayza, D. and T. De Lange, *POT1 as a terminal transducer of TRF1 telomere length control.* Nature, 2003. **423**(6943): p. 1013-8.
18. Kim, S.H., P. Kaminker, and J. Campisi, *TIN2, a new regulator of telomere length in human cells.* Nat Genet, 1999. **23**(4): p. 405-12.
19. Liu, D., et al., *Telosome, a mammalian telomere-associated complex formed by multiple telomeric proteins.* J Biol Chem, 2004. **279**(49): p. 51338-42.

20. Ye, J.Z., et al., *TIN2 binds TRF1 and TRF2 simultaneously and stabilizes the TRF2 complex on telomeres.* J Biol Chem, 2004. **279**(45): p. 47264-71.
21. O'Connor, M.S., et al., *A critical role for TPP1 and TIN2 interaction in high-order telomeric complex assembly.* Proc Natl Acad Sci U S A, 2006. **103**(32): p. 11874-9.
22. Kim, S.H., et al., *TIN2 mediates functions of TRF2 at human telomeres.* J Biol Chem, 2004. **279**(42): p. 43799-804.
23. Li, B., S. Oestreich, and T. de Lange, *Identification of human Rap1: implications for telomere evolution.* Cell, 2000. **101**(5): p. 471-83.
24. de Lange, T., *Shelterin: the protein complex that shapes and safeguards human telomeres.* Genes Dev, 2005. **19**(18): p. 2100-10.
25. Hicke, B.J., et al., *Two versions of the gene encoding the 41-kilodalton subunit of the telomere binding protein of Oxytricha nova.* Proc Natl Acad Sci U S A, 1990. **87**(4): p. 1481-5.
26. Horvath, M.P., et al., *Crystal structure of the Oxytricha nova telomere end binding protein complexed with single strand DNA.* Cell, 1998. **95**(7): p. 963-74.
27. Gray, J.T., et al., *Cloning and expression of genes for the Oxytricha telomere-binding protein: specific subunit interactions in the telomeric complex.* Cell, 1991. **67**(4): p. 807-14.
28. Loayza, D., et al., *DNA binding features of human POT1: a nonamer 5'-TAGGGTTAG-3' minimal binding site, sequence specificity, and internal binding to multimeric sites.* J Biol Chem, 2004. **279**(13): p. 13241-8.
29. Lei, M., et al., *Switching human telomerase on and off with hPOT1 protein in vitro.* J Biol Chem, 2005. **280**(21): p. 20449-56.
30. Fang, G.W. and T.R. Cech, *Molecular cloning of telomere-binding protein genes from Stylonychia mytilis.* Nucleic Acids Res, 1991. **19**(20): p. 5515-8.
31. Fang, G. and T.R. Cech, *Oxytricha telomere-binding protein: DNA-dependent dimerization of the alpha and beta subunits.* Proc Natl Acad Sci U S A, 1993. **90**(13): p. 6056-60.
32. Paeschke, K., et al., *Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo.* Nat Struct Mol Biol, 2005. **12**(10): p. 847-54.
33. Dietmann, S. and L. Holm, *Identification of homology in protein structure classification.* Nat Struct Biol, 2001. **8**(11): p. 953-7.
34. Theobald, D.L., R.M. Mitton-Fry, and D.S. Wuttke, *Nucleic acid recognition by OB-fold proteins.* Annu Rev Biophys Biomol Struct, 2003. **32**: p. 115-33.
35. Holm, L. and C. Sander, *Database algorithm for generating protein backbone and side-chain co-ordinates from a C alpha trace application to model building and detection of co-ordinate errors.* J Mol Biol, 1991. **218**(1): p. 183-94.
36. Chen, J.L. and C.W. Greider, *Determinants in mammalian telomerase RNA that mediate enzyme processivity and cross-species incompatibility.* EMBO J, 2003. **22**(2): p. 304-14.
37. Zaug, A.J., E.R. Podell, and T.R. Cech, *Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro.* Proc Natl Acad Sci U S A, 2005. **102**(31): p. 10864-9.
38. Kuriyan, J. and M. O'Donnell, *Sliding clamps of DNA polymerases.* J Mol Biol, 1993. **234**(4): p. 915-25.

39. Counter, C.M., et al., *Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity.* EMBO J, 1992. **11**(5): p. 1921-9.
40. Teixeira, M.T., et al., *Telomere length homeostasis is achieved via a switch between telomerase- extendible and -nonextendible states.* Cell, 2004. **117**(3): p. 323-35.
41. Brunger, A.T., et al., *Crystallography & NMR system: A new software suite for macromolecular structure determination.* Acta Crystallogr D Biol Crystallogr, 1998. **54**(Pt 5): p. 905-21.
42. Jones, T.A., et al., *Improved methods for building protein models in electron density maps and the location of errors in these models.* Acta Crystallogr A, 1991. **47 (Pt 2)**: p. 110-9.
43. Greider, C.W., *Telomerase is processive.* Mol Cell Biol, 1991. **11**(9): p. 4572-80.
44. Kim, N.W., et al., *Specific association of human telomerase activity with immortal cells and cancer.* Science, 1994. **266**(5193): p. 2011-5.
45. Kim, N.W. and F. Wu, *Advances in quantification and characterization of telomerase activity by the telomeric repeat amplification protocol (TRAP).* Nucleic Acids Res, 1997. **25**(13): p. 2595-7.
46. Ferreira, M.G., K.M. Miller, and J.P. Cooper, *Indecent exposure: when telomeres become uncapped.* Mol Cell, 2004. **13**(1): p. 7-18.
47. Autexier, C. and N.F. Lue, *The structure and function of telomerase reverse transcriptase.* Annu Rev Biochem, 2006. **75**: p. 493-517.
48. Collins, K., *The biogenesis and regulation of telomerase holoenzymes.* Nat Rev Mol Cell Biol, 2006. **7**(7): p. 484-94.
49. Blackburn, E.H., *Telomeres and telomerase: their mechanisms of action and the effects of altering their functions.* FEBS Lett, 2005. **579**(4): p. 859-62.
50. Gao, H., et al., *RPA-like proteins mediate yeast telomere function.* Nat Struct Mol Biol, 2007. **14**(3): p. 208-14.
51. Bertuch, A.A. and V. Lundblad, *The maintenance and masking of chromosome termini.* Curr Opin Cell Biol, 2006. **18**(3): p. 247-53.
52. Hughes, T.R., et al., *The Est3 protein is a subunit of yeast telomerase.* Curr Biol, 2000. **10**(13): p. 809-12.
53. Steinberg-Neifach, O. and N.F. Lue, *Modulation of telomere terminal structure by telomerase components in Candida albicans.* Nucleic Acids Res, 2006. **34**(9): p. 2710-22.
54. Lingner, J., et al., *Three Ever Shorter Telomere (EST) genes are dispensable for in vitro yeast telomerase activity.* Proc Natl Acad Sci U S A, 1997. **94**(21): p. 11190-5.
55. Singh, S.M., et al., *Analysis of telomerase in Candida albicans: potential role in telomere end protection.* Eukaryot Cell, 2002. **1**(6): p. 967-77.
56. Hsu, M., et al., *Mutual dependence of Candida albicans Est1p and Est3p in telomerase assembly and activation.* Eukaryot Cell, 2007. **6**(8): p. 1330-8.
57. Soding, J., A. Biegert, and A.N. Lupas, *The HHpred interactive server for protein homology detection and structure prediction.* Nucleic Acids Res, 2005. **33**(Web Server issue): p. W244-8.

58. Yu, E.Y., et al., *A proposed OB-fold with a protein-interaction surface in Candida albicans telomerase protein Est3.* Nat Struct Mol Biol, 2008. **15**(9): p. 985-9.
59. Friedman, K.L., et al., *N-terminal domain of yeast telomerase reverse transcriptase: recruitment of Est3p to the telomerase complex.* Mol Biol Cell, 2003. **14**(1): p. 1-13.
60. Osterhage, J.L., J.M. Talley, and K.L. Friedman, *Proteasome-dependent degradation of Est1p regulates the cell cycle-restricted assembly of telomerase in Saccharomyces cerevisiae.* Nat Struct Mol Biol, 2006. **13**(8): p. 720-8.
61. Wang, F., et al., *The POT1-TPP1 telomere complex is a telomerase processivity factor.* Nature, 2007. **445**(7127): p. 506-10.
62. Xin, H., et al., *TPP1 is a homologue of ciliate TEBP-beta and interacts with POT1 to recruit telomerase.* Nature, 2007. **445**(7127): p. 559-62.
63. Paeschke, K., et al., *Telomerase recruitment by the telomere end binding protein-beta facilitates G-quadruplex DNA unfolding in ciliates.* Nat Struct Mol Biol, 2008. **15**(6): p. 598-604.
64. Miyoshi, T., et al., *Fission yeast Pot1-Tpp1 protects telomeres and regulates telomere length.* Science, 2008. **320**(5881): p. 1341-4.
65. Do, C.B., et al., *ProbCons: Probabilistic consistency-based multiple sequence alignment.* Genome Res, 2005. **15**(2): p. 330-40.

CHAPTER 3

STRUCTURAL STUDIES OF TAZ1, THE DOUBLE-STRANDED TELOMERIC DNA BINDING PROTEIN IN FISSION YEAST

Taz1 is the double-stranded telomeric DNA binding protein in fission yeast and is proposed to be the homologue of human TRF1 and TRF2. Functionally, Taz1 mimics TRF1 and TRF2 in several aspects. However, there is no structural information of Taz1 to support such proposal. In this chapter, the structural analyses of several domains in Taz1 (TRFH domain, dimerization domain and Rap1-interaction domain) as well as the solution structure of Taz1-Rap1 complex are introduced. Besides, by collaborating with Dr. Julia Cooper's group, the functions of Taz1 dimerization and Taz1-Rap1 complex are also tested *in vivo* and the results are also introduced in this chapter.

3.1 Abstract

Duplex telomeric DNA repeats in fission yeast *Schizosaccharomyces pombe* are bound by Taz1, the proposed homologue of human double-stranded telomeric DNA binding proteins TRF1 and TRF2. By recruiting other yeast telomeric proteins, Taz1 is important for telomere maintenance and regulation. Secondary structural prediction result indicates that Taz1 contains a α-helix-rich region and this region may cover the conserved TRFH domain, which is responsible for dimerization and protein recruitment in TRF1 and

TRF2. By solving the high-resolution crystal structures of TRFH domain ($Taz1_{TRFH}$) and dimerization the domain ($Taz1_D$) in Taz1, we report that $Taz1_{TRFH}$ is structurally different from TRFH domain in TRF1 and TRF2. Rather than using the TRFH domain, Taz1 dimerizes via the formation of a four-helix-bundle. Structural comparison of $Taz1_D$ with the TRFH domains of TRF1 and TRF2 reveals that they employ different architectural principles for homodimerization. The disruption of dimerization interface greatly abolishes Taz1's dsDNA binding ability, as indicated by *in vitro* DNA binding assay and *in vivo* analyses. We also report the solution structure of Taz1-Rap1 complex. Taz1 interacts with Rap1 in a similar way as TRF2-Rap1 complex in human. The recruitment of Rap1 to yeast telomere by Taz1 will greatly regulate telomere length. In summary, Taz1 functionally represents both TRF1 and TRF2 in fission yeast. We conclude that Taz1 is not the structural homologue but the functional homologue of TRF1 and TRF2.

3.2 Introduction

Telomere, the natural end of linear eukaryotic chromosome, is a specialized DNA-protein complex. Telomeric DNA, synthesized by telomerase, contains several hundreds of non-coding repetitive Guanidine-riched DNA sequences that are unique in different organisms (e.g. GGTTAG in vertebrates and $GGTTAC(A)(C)G_{(0-6)}$ in fission yeast), terminating a 3' G-overhang at the very end. Telomere is usually bound with sequence-specific binding telomeric proteins and their interacting partners. At human telomere, six telomeric proteins form a telomere-end-capping complex (shelterin) to protect telomere from degradation, end-to-end fusion and prevent inappropriate DNA repair pathways, such as non-homologous end joining (NHEJ) and homologous recombination (HR). Recently,

several components of a shelterin-like telomere-associated protein complex are also identified in fission yeast *Schizosaccharomyces pombe* [1]. It has been suggested that shelterin complex is evolutionarily conserved from yeast to human. Besides protecting chromosome ends, telomeric protein complex plays an important role in regulating telomerase activity and controlling telomere length.

Taz1 was identified by the one-hybrid screening in fission yeast *Schizosaccharomyces pombe* and it has been proposed to be the homologue of human double-stranded (ds) telomeric DNA binding proteins TRF1 and TRF2[2-4]. Taz1 stays on yeast telomeric DNA track and is important in yeast telomere end protection and telomere length regulation. Deletion of *taz1* gene leads to the release of single-stranded 3' G-overhang regulatory machinery, massive elongation of double-stranded telomeric repeats, and eventually causes the loss of telomeric chromatin structure and de-repression of telomere-adjacent transcription [2]. Taz1 is also involved in 3' overhang processing and protecting yeast cells from cold sensitivity [5, 6]. Several proteins have been found to interact with Taz1 directly, including yeast telomeric proteins Rif1 and Rap1 [7, 8]. By recruiting Rif1 and Rap1 to telomere, Taz1 negatively regulates telomere length through different independent pathways [6, 7].

Taz1 highly resembles TRF1 and TRF2 in several aspects. Taz1 is able to count the telomeric repeats, and thus function in the similar "protein-counting" model as budding yeast Rap1 and TRF1 [9-11]. Both Taz1 and TRF1 are important for gene transcription regulation. Taz1 recruits Rap1 to yeast telomere region, in a similar way as TRF2 brings Rap1 to human telomere. Functional assay results already show that similar to TRF1 and TRF2, Taz1 dimerizes on the telomere track [12]. Based on sequence

alignment with TRF1 and TRF2, Taz1 contains the highly conserved Myb domain in the C-terminus. The Myb domain, shared by all the telomeric dsDNA binding proteins, is responsible for mediating dsDNA-binding activity. In budding yeast *S. cerevisiae*, Rap1 binds to dsDNA via its two tandem Myb domains. Human TRF1 and TRF2 only contain one Myb domain per molecule. In order to arrange two Myb domains for dsDNA recognition as *Sc*Rap1 does, TRF1 and TRF2 homodimerize via the central TRFH (TRF Homology) domain [4]. Crystal structures of TRFH domain in TRF1 and TRF2 indicate that this α-helix-rich domain is responsible for dimerization [4]. Mutations that disrupt the dimerization interface of TRFH domain will greatly reduce TRF1/TRF2's binding affinity to dsDNA, as suggested by *in vitro* dsDNA binding assay. So far, TRFH-like proteins have been found not only in human, but also in some simple organisms such as *Trypanosoma* [13]. It even has been proposed that besides Myb-domain, TRFH domain is another signature motif that is conserved through all the dsDNA binding telomeric proteins. This hypothesis is supported by the discovery of a 270-aa region locating in the center of Taz1. Secondary structural prediction indicates that this region is made up of α-helices and the helix arrangement is very close to TRF1/TRF2-TRFH domain. Therefore, it was suspected that this region is the TRFH domain of Taz1 and that it may mediate Taz1 homodimerization. However, there has been no structural evidence to support such hypothesis.

We were very curious about the structure of Taz1. Specifically, we want to test whether that 270-aa-long region is indeed the Taz1 TRFH domain and is responsible for dimerization. We are also interested in the Taz1-Rap1 complex since our recent discovery indicates that human TRF2-Rap1 complex plays an essential role in telomere end

protection (Chen Y et al., 2009, submitted). In order to address these questions, structural studies on the potential TRFH domain in Taz1 were preformed. The solution structure of the interacting domains of Taz1-Rap1 complex was determined by NMR. Based on this structural information, the biological relevance of Taz1 dimerization and the function of Taz1-Rap1 complex were also tested *in vivo*.

3.3 Structural determination of the TRFH domain of Taz1

Sequence alignment and secondary structure prediction analyses suggested that Taz1 contains a putative TRFH domain (residues 117 – 391) [3]. However, $Taz1_{117\text{-}391}$ only showed a low sequence similarity to the TRFH domains of human TRF1 and TRF2, and the size of $Taz1_{117\text{-}391}$ was significantly larger than those of $TRF1_{TRFH}$ and $TRF2_{TRFH}$ (~ 200 amino acids), resulting in big gaps in the sequence alignment [14]. In addition, unlike $TRF1_{TRFH}$ and $TRF2_{TRFH}$ that form stable homodimmers, purified $Taz1_{117\text{-}391}$ is a monomor in solution as revealed by gel filtration chromatographic analysis (Fig. 3.1a). Thus, the decisive answer to the similarity between $Taz1_{117\text{-}391}$ and the TRFH domains of TRF proteins required a structural approach. $Taz1_{113\text{-}400}$ was cloned into the SUMO-His_6-pET28 vector and the protein was overexpressed and purified to homogeneity. Intensive crystallization screening and optimization only yielded twisted needle crystals of $Taz1_{113\text{-}400}$ that were not suitable for diffraction. Therefore, $Taz1_{113\text{-}400}$ was subjected to limited proteolysis analysis using four different proteases (elastase, Glu-C, trypsin, and subtilisin) (Fig. 3.1b). A stable digestion product was obtained after 90-minute treatment of trypsin (Fig. 3.1c). The stable fragment was identified by mass spectrometry to contain residues 127-388 of Taz1 (Fig. 3.1d). $Taz1_{127\text{-}388}$ was then expressed, purified and

successfully crystallized in space group $P4_12_12$ with one molecule per asymmetric unit (Fig. 3.2a-c) (Table 3.1). The protein was phased to 2.7 Å resolution by the multi-wavelength anomalous dispersion (MAD) method by using selenium-methionine substituted crystals, and the phases were extended with the native data set to 2.3 Å resolution (Table 3.1). In the refined structure, C-terminal residues 362-388 were not represented in the electron density and therefore we presume that this region is disordered in solution. Hereafter, we will refer to $Taz1_{127\text{-}361}$ as the TRFH domain of Taz1, $Taz1_{TRFH}$.

3.4 Crystal structure of $Taz1_{TRFH}$

$Taz1_{TRFH}$ adopts a compact globular fold, with 14 α-helices tightly packed around a central hydrophobic core (Fig. 3.3a). The helices are organized so that the entire domain resembles the shape of a slightly curved hockey puck with a diameter of ~ 50 Å (Fig. 3.3b). Of the 14 helices, α1B, α3, α6, α7, α8, α9, α12, and α13 form an eight-helix bundle, while the other six short helices α1A, α2, α4, α5, α10, and α11, cap the exposed hydrophobic surface from the sides of the central helical bundle (Fig. 3.3a). The eight-helix bundle can be further divided in to four groups of helix pairs (α1B and α3, α6 and α13, α7 and α8, α9 and α12). Hydrophobic contacts between the helices within each pairs are the main driving force to maintain the central helical core. For instance, the aromatic side chains of three phenylalanine residues (Phe263 from α8, and Phe350 and Phe354 from α13) stack with one another to help holding helix α8 and α13 together (Fig. 3.3c). In contrast to the hydrophobic nature of the central core, an analysis of surface electrostatic potential reveals a large stripe of strongly electronegatively charged surface on the convex side of the structure (Fig. 3.3d), consisting of a cluster of 13 acidic

residues that are distributed in several peripheral short helices (Fig. 3.3e). This negatively charged surface can potentially mediate protein-protein interactions with other telomeric factors. Further investigation on Taz1 and its binding partners is needed to test this hypothesis.

3.5 Comparison of $Taz1_{TRFH}$ with the TRFH domains of TRF1 and TRF2

$Taz1_{TRFH}$ and the TRFH domains of TRF1 and TRF2 are all α-helical structures. The topology of the helical core of $Taz1_{TRFH}$ is similar to $TRF1_{TRFH}$ and $TRF2_{TRFH}$ (Fig. 3.4a, d). Despite this topology similarity, $Taz1_{TRFH}$ is not the structural homolog of the TRFH domains of TRF1 and TRF2. There are significant structural differences between $Taz1_{TRFH}$ and the TRFH domains of TRF1 and TRF2 (Fig. 3.4a, d). Only six of 14 $Taz1_{TRFH}$ helices can be matched in the TRF1/2 TRFH domains and the rest eight helices in $Taz1_{TRFH}$ cannot be superimposed with $TRF1_{TRFH}$ and $TRF2_{TRFH}$. Consequently, the root-mean-square deviation (r.m.s.d.) between $Taz1_{TRFH}$ and the TRFH domains of TRF1 and TRF2 is over 4.5 Å, suggesting that $Taz1_{TRFH}$ and the TRF1/2 TRFH domains are not structurally closely related. Consistent with this notion, structural based sequence alignment show a less than 6 % identity between $TRF1_{TRFH}$ and the TRFH domains of TRF1 and TRF2 with many large gaps (Fig. 3.4d).

Notably, as indicated by biochemical analysis (Fig. 3.1a, Fig. 3.2a), the most prominent difference between $Taz1_{TRFH}$ and the TRFH domains of TRF1 and TRF2 is that $Taz1_{TRFH}$ adopts a monomeric conformation, whereas TRF1/2 TRFH domains are homodimers. In both $TRF1_{TRFH}$ and $TRF2_{TRFH}$, the α1 helix is only slightly bent (< 5°) and retains its integrity as a single α-helix through maintenance of all intra-helical

hydrogen bonds (Fig. 3.4a). The N-terminal portion of α1 extends outside of the helical core and interacts with helix α10 from the other molecule in the dimer (Fig. 3.4b). In contrast, the α1 helix of $Taz1_{TRFH}$ is severely bent (~ 70°), and is stretched between residues 138 and 140, resulting in the disruption of hydrogen bonds between the carbonyl oxygen atoms of residues 138 and 139, and the amide nitrogen atoms of residues 142 and 143, respectively. Consequently, the α1 helix in $Taz1_{TRFH}$ is now composed of two helices α1A and α1B, with a one-residue intervening linker (Fig. 3.3a, Fig. 3.4a). The bended α1A helix and the N-terminal tail of $Taz1_{TRFH}$ fits into a hydrophobic groove formed by helices α3, α4, and α5 (Fig. 3.4c). This conformation is further stabilized by a pair of hydrogen bonding interactions between main-chain atoms of Trp129 and the hydroxyl groups of Tyr206 and Tyr213, respectively (Fig. 3.4c). Because of the bending, $Taz1_{TRFH}$ lacks the dimeric interface mediated by helices α1 and α10 in the structures of the TRFH domains of TRF1 and TRF2 [4]. In addition, the C-terminal tail of $Taz1_{TRFH}$ (residues 362-388) that corresponds to helix α10 in TRF1 and TRF2 is also disordered in the $Taz1_{TRFH}$ structure. Thus, the structural features of both the bending of α1 and the lack of α10 explain why $Taz1_{TRFH}$ adopts a monomeric conformation.

3.6 Identification and structure determination of the dimerization domain of Taz1

Both our structural and biochemical analyses reveal that, unlike the TRFH domain of human TRF1 and TRF2, $Taz1_{TRFH}$ adopts a monomeric conformation (Fig. 3.1a, Fig 3.2a, Fig. 3.3a). However, previous studies indicated that Taz1 binds to telomeric DNAs as a homodimer [12], suggesting that the dimerization of Taz1 must be mediated by another region outside of the TRFH domain. To identify the dimerization domain of Taz1,

various fragments were evaluated for their ability to form a homodimer in solution by gel filtration chromatography and chemical cross-linking analysis (Fig. 3.5a). We found that a fragment of Taz1 containing residues 395 – 490 of Taz1 constitutes the minimal structural core that is necessary and sufficient for dimerization (Fig. 3.5a).

To understand the mechanism of Taz1 dimerization, we reconstituted and crystallized the $Taz1_{395-490}$. We determined the crystal structure at a resolution of 1.5 Å by single isomorphous replacement (SIR) using an iodide compound (NaI) (Table 3.2). The final model contains residues 409 – 478 of Taz1, and has been refined to an R-value of 25.40% (R_{free} = 26.95%). No electron density is observed corresponding to the N-terminal residues 395 – 408 or the C-terminal residues 479 – 490, and we presumed that these two regions are disordered in solution. Hereafter we refer to $Taz1_{409-478}$ as the dimerization domain of Taz1 ($Taz1_D$) (Fig. 3.5b-e).

3.7 Crystal structure of the dimerization domain of Taz1

In the crystals, two molecules of $Taz1_D$ form a homodimer, resulting the burial of a total of 2514.96 $Å^2$ of the solvent accessible surface area (Fig. 3.6a, c). Each $Taz1_D$ monomer consists of three helices. Helices Dα2 and Dα3 pack closely against each other forming an antiparallel coiled-coil that dimerizes through a two-fold axis of crystallographic symmetry (Fig. 3.6b). The symmetry related Dα2 helices pack against each other parallelly while the two Dα3 helices open up at the C-terminal end resulting in a slightly deformed four-helix bundle (Fig. 3.6b). The N-terminal helix of $Taz1_D$, Dα1, contains only two helical turns and packs perpendicularly on Dα2 and Dα3 opposite the dimeric interface (Fig. 3.6a, b). Dα1 is linked with αD2 through an eleven-residue loop, which

together with helix αD1 is attached to the coiled-coil through van der Waals contacts. Particularly, the side chains of L415, L417, I420, and I425 insert into a hydrophobic groove formed by residues from helices Dα2 and Dα3 (Fig. 3.6d). The position of αD1 is further stabilized by two hydrogen-bonding interactions; the hydroxyl group of S411 and the main chain carbonyl oxygen of L415 on Dα1 accept two hydrogen bonds from the side chain amino groups of Q456 and K440, respectively (Fig. 3.6d).

The hydrophobic packing contact at the dimeric interface of $Taz1_D$ is extensive, with 12 residues from one molecule interacting with those from the other molecule. The interface consists of seven planes of twofold symmetry related interdigitating residues from helices Dα2 and Dα3 (Fig. 3.7a). Hydrophobic residues stack closely against each other both within and between adjacent planes. Within each plane, the interface is contributed by the residues at positions *a* and *d* of the heptad repeats, a characteristic of coiled-coil proteins (reviewed in [15-17]). Although the dimeric interface is predominantly hydrophobic, intermolecular electrostatic interactions provide additional specificity and stability to the structure. At the end of the helix bundle, the side chain of Arg471 makes two salt bridge interactions with the carboxyl group of Glu427 from the other molecule, and the aliphatic portions of the side chains of the two symmetry-related Glu427 stack with each other connecting the two electrostatic interaction networks together (Fig. 3.7a, b). These intermolecular interactions are further buttressed by three hydrogen bonds formed by Arg471 and the main-chain carbonyl groups of Ile420, Asn422 and Ile425 (Fig. 3.7a, b). At the center of the helix bundle, Ser461 at position *a* in plane 3 accepts a hydrogen bond from the side chain hydroxyl group of Thr442 while donates another one to the main chain carbonyl of Leu438 from plane 4 (Fig. 3.7a, b).

Structural comparison of $Taz1_D$ with the TRFH domains of TRF1 and TRF2 reveals that they employ different architectural principles for homodimerization. In TRF1 and TRF2, helices α1 and α2 from both monomers interact with each other to generate an anti-parallel four-helix bundle with the symmetry-related dyad perpendicular to the helical axes (Fig. 3.4b). The third helix α10 at the C-terminus of the TRFH domain protrudes outside of the TRFH domain and packs with the N-terminal portion of helix α1 forming a cross-brace to stabilize the four-helix bundle (Fig. 3.4b). In contrast, although $Taz1_D$ also consists three helices, its dimer interface only involves helices Dα1 and Dα2, which form a parallel four-helix bundle with the symmetry-related dyad parallel to the helical axes (Fig. 3.6a, b, Fig. 3.7a, b).

3.8 Biochemical and functional analyses of the Taz1 dimerization interface

To corroborate our structural analysis, we assessed the contribution of the residues at the $Taz1_D$ dimeric interface to the stability of Taz1 dimer structure. Four single point mutations, L431R, V434W, L438W, and L445R were generated and tested in a yeast two-hybrid assay. All these residues are in the center of the hydrophobic dimer interface, and we predicted that substitution of them with charged residue arginine or large hydrophobic residue tryptophan would disrupt the packing within the dimer structure. Indeed, compared with the wild-type Taz1, three mutants (L431R, V434W, and L445R) completely abolished the dimeric interaction of Taz1 (Fig. 3.8a). Notably, tryptophan substitution of Leu438 still maintained a detectable dimeric interaction (~ 10% of the wild type level), consistent with the observation that the binding pocket for Leu438 in the

crystal structure is deeper than that for Val434 (Fig. 3.8b). These results confirmed that the interface observed in the $Taz1_D$ crystal structure is responsible for Taz1 dimerization.

A common feature of the telomeric dsDNA binding proteins is that they all bind to the duplex DNA via multiple Myb domains. In budding yeast *S. cerevisiae*, Rap1 contains two tandem Myb domains. Whereas in humans, although TRF1 and TRF2 contain only one Myb domain, their DNA binding activities require formation of homodimers *in vivo* and *in vitro* [4, 18, 19]. Based on the facts that Taz1 also exists as a homodimer [12](Fig. 3.6) and that each monomer contains only one Myb domain, we hypothesized that Taz1 homodimerization is also essential for its telomere association. To test this idea, the L445R mutant that disrupted the Taz1 dimerization in the yeast two-hybrid assay was studied for its effect on the DNA binding activity by an electrophoretic mobility shift assay (EMSA) with a duplex $(GGTTACA)_3$ probe. Because attempts to express recombinant full-length Taz1 yielded insoluble protein, a C-terminal fragment that includes both the dimerization and the Myb domains ($Taz1_{D\text{-}Myb}$, residues 408–663) was instead used in the assay. To determine the equilibrium dissociation constant (*K*d) of the $Taz1_{D\text{-}Myb}$–DNA complex, the binding of a trace amount of DNA probe with increasing $Taz1_{D\text{-}Myb}$ was measured. Wild-type $Taz1_{D\text{-}Myb}$ bound to DNA in a simple hyperbolic binding equilibrium with a *K*d of ~ 600 nM (Fig. 3.9a). In contrast, the substitution of L445 with an arginine reside caused a 10-fold decrease in DNA binding with a *K*d of ~ 7 μM (Fig. 3.9b). Notably, the Myb domain of Taz1 ($Taz1_{Myb}$) exhibits a similar level of reduction in DNA binding affinity (*K*d ~ 60 μM) (data not shown). Collectively, we conclude that similar to TRF1 and TRF2, homodimerization of Taz1 is required for its efficient association with the telomeric duplex DNA.

To determine the *in vivo* roles of Taz1 dimerization in telomere maintenance, we examined the telomere length in yeast cells expressing wild-type Taz1 versus Taz1 mutants that are deficient in dimerization (a collaboration with Dr. Julia P. Cooper at Cancer Research UK). Consistent with the published results, deletion of *taz1*$^{+}$ from yeast cells resulted in dramatic increase in telomere length and length heterogeneity compared to wild-type cells (Fig. 3.10a). Notably, three Taz1 single mutations (L431R, V434W, and L445R) that showed no detectable dimerization ability in yeast two-hybrid assay exhibited significant loss of function in telomere length regulation and manifested phenotypes that were as severe as that is in *Taz1Δ* cells (Fig. 3.10a, lane 3, 4 and 6). Intriguingly, the L438W mutant that retained partial dimerization ability in the yeast two-hybrid assay also retained partial function in suppressing telomere elongation (Fig. 3.10, lane 5). Thus, these data suggested that Taz1 dimerization is critical for telomere length regulation.

3.9 Characterization of the Taz1-Rap1 interaction

Fission yeast Rap1 is recruited to telomeres by binding to Taz1. Both Taz1 and Rap1 are involved in regulating telomere length homeostasis, 3' G-overhang processing and preventing NHEJ-medicated chromosome end-to-end fusion [6]. Like budding yeast and mammalian Rap1 proteins, *S. pombe* Rap1 contains a N-terminal BRCT (BRCA-1 carboxyl-terminal) domain and a central Myb domain [7, 8]. However, the RCT (Rap1 carboxyl-terminal) domain in pombe Rap1 is not similar as those in *Sc*Rap1 and huma Rap1 according to the sequence alignments. The RCT domain of mammalian RAP1 mediates the interaction with TRF2. Thus, it was unknown how *S. pombe* Rap1 is

recognized and recruited to the chromosome ends by Taz1.

To determine the mechanism of Rap1 recognition by Taz1, we characterized the Taz1-Rap1 interaction (Fig. 3.11a). Various fragments of the GST-Rap1 protein were evaluated for their ability to interact with Taz1. The N-terminal half (residues 1–570) of Rap1 did not show detectable binding to Taz1. In contrast, the C-terminal half (residues 567–693) formed a stable complex with Taz1 (data not shown). Further mapping revealed that, similar to the mammalian TRF2-RAP1 interaction, the C-terminus of Rap1 (resides 639-693) is sufficient for interaction with Taz1 (Fig 3.11b). Using a similar strategy, a 32-residue fragment of Taz1 immediately after the TRFH domain (residues 365-396) was found to be the minimal structural core that is necessary and sufficient for binding to Rap1. Hereafter, $Taz1_{365\text{-}396}$ and $Rap1_{639\text{-}693}$ will be referred to as $Taz1_{RBD}$ (Rap1-binding domain) and $Rap1_{RCT}$, respectively (Fig. 3.11c).

3.10 Solution structure of the $Taz1_{RBD}$-$Rap1_{RCT}$ complex

To reveal the structural basis of Rap1 recognition by Taz1, we reconstituted the $Taz1_{RBD}$-$Rap1_{RCT}$ complex. However, after extensive constructs and crystallization screening, we failed to obtain crystals suitable for structural study. Based on the facts that the $Taz1_{RBD}$-$Rap1_{RCT}$ complex is very stable at 25°C and that the size of the complex is ~ 9.6 kDa that falls well into the feasible range for NMR technique, we set up a collaboration with Dr. Hongyu Hu's laboratory (Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, Shanghai, China) to determine the solution structure of the $Taz1_{RBD}$-$Rap1_{RCT}$ complex by nuclear magnetic resonance (NMR). To simplify the ^{15}N- and ^{13}C-labeled NMR sample preparation, we linked $Rap1_{RCT}$ to $Taz1_{RBD}$ with a 14-residue linker, which

contains four Gly-Gly-Ser repeats with one Hind III restriction enzyme site in the middle (Fig. 3.11c). The linker is flexible and long enough so that it does not influence the proper interaction between $Rap1_{RCT}$ and $Taz1_{RBD.}$ The high-resolution structure of the $Taz1_{RBD}$-$Rap1_{RCT}$ complex (Fig. 3.12) was calculated from a total of 2420 NMR restrains (Table 3.3).

The solution structure of the $Taz1_{RBD}$-$Rap1_{RCT}$ complex reveals a compact globular fold, except for the linker region that is unstructured. $Taz1_{RBD}$ contains a single helix αTaz1, while $RAP1_{RCT}$ consists of three α helices (Fig. 3.12a, b). Together, these helices are arranged into an intermolecular four-helix bundle. The helix from $Taz1_{RBD}$ packs against helices α1 and α2 of $RAP1_{RCT}$ to form an intermolecular three-helix bundle (Fig. 3.12a, b). The formation of the binary complex involves an extensive set of interactions and causes the burial of 1,682 $Å^2$ of surface area at the interface. The driving force for the binding of Taz1 to Rap1 is van der Waals interactions (Fig. 3.12c, d). Four hydrophobic residues of $Taz1_{RBD}$ (Ile379, Leu380, Leu383 and Val387) make extensive contacts with the hydrophobic wedge between helices α1 and α2 of $Rap1_{RCT}$ (Fig. 3.12c, d). Located at the center of the hydrophobic interface, the side chain of Ile379 and Leu383 of Taz1 are deeply buried in a hydrophobic groove formed by a group of hydrophobic residues of Rap1 (Fig. 3.12e). In addition to hydrophobic contacts, electrostatic interactions also contribute to the formation of the $Taz1_{RBD}$-$Rap1_{RCT}$ complex. On one side of the Taz1 helix, the guanidinium group of Arg386 of Taz1 is directed towards a negative charged depression of Rap1 and anchored in place through a network of electrostatic interactions: three salt bridges and one hydrogen bond to the carboxyl groups of three residues of Rap1 (Asp652 and Asp656 from α1 and Glu666

from α2) (Fig. 3.12c). On the other side the Taz1 helix, another arginine residue of Taz1, Arg384, makes two direct hydrogen bonds to the side chain of Rap1 Glu674 (Fig. 3.12c). Collectively, these electrostatic contacts function as the two arms of a clamp to help $Rap1_{RCT}$ hold the $Taz1_{RBD}$ helix.

The structure of $Rap1_{RCT}$ closely resembles the N-terminal three-helix bundle of the RCT domains of both budding yeast *Sc*Rap1 and human hRAP1 even with no apparent sequence similarity (Fig. 3.13a). An unbiased search for structurally homologous proteins using the Dali server [20] revealed that the structure of $Rap1_{RCT}$ is most similar to that of human $RAP1_{RCT}$; the two RCT domains can be superimposed with an rmsd of 2.7 Å for 104 equivalent C_α pairs (Fig. 3.13b). In addition to the structurally conserved helix bundle, the RCT domains of *Sc*Rap1 and hRAP1 contain an additional three-helix bundle at the C-termini of the proteins (Fig. 3.13b). Notably, the sequence of this extra helix bundle is highly conserved in budding yeast and mammalian Rap1 proteins but is not present in fission yeast Rap1 (Fig. 3.13a). This explains the failure to detect the RCT domain of *S. pombe* Rap1 by bioinformatics. The structural similarity does not only limit to the $Rap1_{RCT}$ moiety of the complex. Strikingly, when the structures of the $Taz1_{RBD}$-Rap_{RCT} and the human $TRF2_{RBD}$-$RAP1_{RCT}$ complexes were superposed based on the RCT domains of Rap1 proteins, it is evident that the helix αA of $Taz1_{RBD}$ packs against the hydrophobic groove of $Rap1_{RCT}$ in a very similar fashion as the α1 helix of $TRF2_{RBD}$ in the $TRF2_{RBD}$-$Rap1_{RCT}$ complex (Fig. 3.13b). $Taz1_{RBD}$ lacks the second helix α2 in $TRF2_{RBM}$, which makes less contribution to the $TRF2_{RBD}$-$Rap1_{RCT}$ interaction (Fig. 3.13b). Taken together, these structural similarities strongly support the

notion that Taz1 and Rap1 are the fission yeast homologues of human TRF2 and RAP1, respectively.

3.11 Biochemical and functional analyses of the Taz1-Rap1 interface

To corroborate our structural analysis, we examined whether missense mutations of the interface residues of $Taz1_{RBD}$ or $Rap1_{RCT}$ could weaken or disrupt the Taz1-Rap1 interaction. Consistent with the crystal structure, substitution of Taz1 Ile379 and Leu383 with a positively charged and bulkier arginine residue completely abolished the interaction with Rap1 in a yeast two-hybrid assay (Fig. 3.14a). Similarly, Rap1 mutations I655R on the other side of the interface also impaired the interaction (Fig. 3.14a). These results indicated that a single point mutation at the hydrophobic interface is sufficient to disrupt the Taz1-Rap1 interaction. In contrast, two point mutations ($Taz1_{L380R}$ and $Rap1_{V651R}$) only partially disrupted the interaction and another one $Taz1_{V387R}$ maintained a wild-type binding with Rap1 (Fig. 3.14a), suggesting these residues are not essential for binding.

To further examine the *in vivo* roles of the Taz1-Rap1 interaction in telomere length regulation, we analyzed the phenotypes of four Taz1 mutations ($Taz1_{I379R}$, $Taz1_{L380R}$, $Taz1_{L383R}$, and $Taz1_{L387R}$) that showed varying degrees in disruption of the Taz1-Rap1 interaction in the yeast two-hybrid assay. Notably, all four point mutants of Taz1 exhibited partial to complete loss of telomere length regulation, in a manner that is entirely consistent with the severity of the Rap1-binding defect (Figs. 3.14b, lane 4-7). The two mutants ($Taz1_{I379R}$ and $Taz1_{L383R}$) that completely lost the Rap1-binding activity in the yeast two-hybrid assay showed a *rap1Δ*-like telomere length defect (Fig. 3.14b,

lane 4, 6). Furthermore, the $Taz1_{L387R}$ mutant that retained almost wide-type Rap1-binding ability in the yeast two-hybrid assay also exhibited the mildest defect in suppressing telomere length elongation (Fig. 3.14b, lane 7). Taken together, our mutagenesis analyses suggest that the hydrophobic interface is necessary for both *in vitro* and *in vivo* binding of Taz1 to Rap1.

3.12 Discussion

As one important component of yeast telomere-associated protein complex, Taz1 is previously thought to mimic its human homologues TRF1 and TRF2 in many aspects, as supported by the following evidence. 1.) By sequence alignment and secondary structural prediction, Taz1 contains the conserved C-terminal Myb-domain, which is the feature motif of all the known double-stranded telomeric DNA binding proteins. The TRFH domain is also considered to locate in the center of Taz1, although the sequence identity and similarity is low compared with TRFH domain in TRF1 and TRF2 [4]. 2.) Similar to TRF1 and TRF2, Taz1 functions as a homodimer on the fission yeast telomeric dsDNA [12]. Through direct interaction, Taz1 recruits Rap1 to telomere, reassembles the human TRF2-Rap1 complex in fission yeast [7, 8]. Before our structural results are reported, it has been widely proposed that Taz1 homodimerizes through its TRFH domain, which shares the similar structural architecture as TRF1 and TRF2. However, to our great surprise, Taz1 adopts a novel method for dimerization. Due to the different helix arrangements, $Taz1_{TRFH}$ forms a more compact helix-bundle and lacks the two extended α helices (α1 and α10 in TRF1/TRF2-TRFH) that are required for the cross-brace-like homodimerization interface in TRF1 and TRF2. Alternatively, Taz1 dimerizes via two α

helices outside $Taz1_{TRFH}$, which are paralleled packed to form a four-helix-bundle with the symmetry-related dyad parallel to the helical axes.

So far, the dsDNA binding telomeric proteins have been found in many organisms, from the simple *Trypanosoma* [13] to yeast and higher vertebrates. All the functional evidences suggest that they are very crucial in telomeric pathways, such as telomere end protection, proper telomere structure maintenance, as well as facilitating efficient telomere replication. Although Myb domain is shared by all the known dsDNA binding telomeric proteins, the TRF-like complex, rather than Myb domain, is previously considered to represent the ancestral scenario. In fission yeast and higher vertebrates, duplex telomeric DNA repeats are bound by TRFH containing proteins and Rap1 is directed to telomere via interactions with TRF homologues. In budding yeast, Rap1 orthologue stays on telomeric DNA track and the TRF homologue is missing. This exception may be caused by a duplication of Myb domain gene during the revolution. *Sc*Rap1 accidentally obtained two Myb domains, which made it possible to bind duplex telomeric DNA by itself. Meanwhile, the TRF homologue got lost for some unknown reasons. Therefore, before our structural studies on Taz1 TRFH domain, it has been hypothesized that TRFH domains in all the TRF homologues bear the same structural architecture and mediate dimerization. However, our crystal structures of $Taz1_{TRFH}$ and $Taz1_{D}$ challenge this hypothesis. Future structural studies of the potential TRF homologues in other species will help further clarify the debate.

Nevertheless, despite the structural differences in TRFH domain, we still consider Taz1 as the homologue of TRF1 and TRF2 because of the common functions. Being the only telomeric protein that binds yeast dsDNA, Taz1 carries out several biological

functions that have been distinctively distributed to two players in human: TRF1 and TRF2. For example, both Taz1 and TRF1 are actively involved in efficient replication of telomeric repeats [21, 22]. The N-terminal domains of both proteins are very acidic (pI_{TRF1N}=3.5, pI_{Taz1N}=4.8), although the function is still unknown [23]. On the other hand, Taz1 provides a landing platform for other yeast telomeric proteins, just similar as TRF2 does in human. Our solution structure of Taz1-Rap1 complex clearly reveals that the TRF2-Rap1 complex is similarly reassembled in fission yeast. In the Taz1-Rap1 complex, Taz1 shares the similar α helix as that in TRF2, which contributes most to the Rap1-interacting interface. *In vivo* analyses also indicate that the recruiting of Rap1 to yeast telomere is important for telomere length regulation, which match our recent discoveries about TRF2-Rap1's function on human telomere protection (Chen, Y et. al, 2009, submitted) Furthermore, although Taz1 dimerizes differently from TRF1 and TRF2, the purpose of the dimerization remains the same: to assemble two Myb domains that are required for binding to dsDNA with high affinity, as suggested by our EMSA data. Collectively, we draw the conclusion that rather than a structural homologue, Taz1 is the functional homologue of TRF1 and TRF2. Taz1 represents both TRF1 and TRF2 in fission yeast and plays an important role in the telomere length regulatory mechanism.

3.13 Methods and materials

Protein expression and purification

The TRFH domain of Taz1 (residues 128 – 388) and dimerization domain of Taz1 (residues 395 – 490) were cloned into a modified pET28b vector with a SUMO protein fused at the amino-terminus after the His_6 tag [24]. The TRFH domain of Taz1 was

expressed in *E. coli.* BL21(DE3). After induction for 16 hours with 0.1 mM IPTG at 25 °C, the cells were harvested by centrifugation and the pellets were resuspended in lysis buffer (50 mM Tris-HCl pH 8.0, 50 mM NaH_2PO_4, 400 mM NaCl, 3 mM imidazole, 10% glycerol, 1mM PMSF, 0.1 mg/ml lysozyme, 2 mM 2-mercaptoethanol, and home-made protease inhibitor cocktail). The cells were then lysed by sonication and the cell debris was removed by ultracentrifugation. The supernatant was mixed with Ni-NTA agarose beads (Qiagen) and rocked for 6 hours at 4 °C before elution with 250 mM imidazole. The ULP1 protease was added to remove the His_6-SUMO tag. Finally the $Taz1^{TRFH}$ protein was further purified by passage through Mono-Q ion-exchange column and by gel-filtration chromatography on Hiload Superdex 75 equilibrated with 25 mM Tris-HCl pH 8.0, 150 mM NaCl and 5 mM dithiothreitol (DTT). The purified $Taz1^{TRFH}$ protein was concentrated to 25 mg/ml and stored at -80 °C. The Seleno-Met substituted $Taz1^{TRFH}$ protein was similarly purified.

The protein of dimerization domain of Taz1 was expressed in *E. coli.* and purified following the same procedures as described above.

The $Rap1_{RCT}$ and $Taz1_{RBD}$ were artificially linked together with a 14-residue linker, which contains 4 Gly-Gly-Ser repeats with one Hind III restriction enzyme site in the middle. The fusion protein of $Rap1_{RCT}$-$Taz1_{RBD}$ was expressed in *E. coli.* and purified following the same procedures as described above.

Expression and Purification of the Fusion Protein $Taz1_{RBD}$-$Rap1_{RCT}$ (spRT6)

The spRT6 construct was expressed in *E. coli* BL21(DE3) strain. 10 mL of overnight culture was transferred to 1 L of M9 minimal media supplemented with $^{15}NH_4Cl$ or

$^{15}NH_4Cl$/[^{13}C]-glucose for the preparation of ^{15}N-labeled or $^{15}N/^{13}C$-labeled proteins, respectively. When the culture reached an OD_{600} of 0.6, the expression of spRT6 fusion protein was induced by adding IPTG to a final concentration of 0.2 mM. The proteins were purified by Ni-NTA affinity chromatography. The SUMO-tag was removed by on-column cleavage with ULP1. The proteins were further purified by gel filtration on a Superdex-75 column (Amersham Biosciences).

Limited protease (Trypsin) cleavage of $Taz1_{113-400}$

The protein of $Taz1_{113-400}$ was incubated with 0.2% w/w Trypsin (Roche) at 25 °C in 25 mM Tris-HCl pH 8.0, 150 mM NaCl, and 5 mM DTT. At various time points, 10 μl aliquots of the reaction mixture were withdrawn, diluted with 10 μl of water and 5 μl of SDS loading dye, and run on 15% SDS-PAGE visualized with Coomassie brilliant blue stain.

MALDI mass spectrometry of the limited protease (Trypsin) cleavage products

Fro MALDI mass spectrometry analysis, $Taz1_{113-400}$ was incubated with 0.2 % w/w Trypsin (Roche) at 25 °C in 25 mM Tris-HCl pH 8.0, 150 mM NaCl, and 5 mM DTT. Aliquots were withdrawn as described above for SDS-PAGE analysis. At the 90 min time point, 2 μl of the reaction mixture was co-crystallized with 2 μl sinapinic acid matrix. The samples were analyzed by MALDI-TOF-MS in linear mode. The major product by MALDI had an MH(+1) of 29360.2 $\pm$ 30 Da. Examination of the map of predicted Trypsin sites revealed that this fragment corresponds to the predicted fragments: $Taz1_{127-383}$ [MH(+1) 29394.6 Da].

Electrophoretic mobility shift assay (EMSA)

The protein of $Taz1_{408-663}$ in binding buffer (25 mM Tris-HCl pH 8.0, 150 mM NaCl, and 5 mM DTT) was mixed with 1 μM 32P-labeled telomeric dsDNAs (32mer) in a total volume of 20 μL. The reaction mixtures were incubated at 4 °C for 30 min. Then the mixtures were directly loaded onto a 4−20% nondenaturing polyacrylamide gel. Electrophoresis was carried out in TBE buffer at 150 V for 85 min at 4 °C. The gels were dried, and radiolabeled dsDNA was visualized using a PhosphorImager. (32mer dsDNA was annealed by the following ssDNA with its complementary strand: 5' - GATCTCAGCT GGTTACA GGTTACA GGTTACA G- 3') The signal of the radiolabeled dsDNA was quantified by using ImageQuant (GE Healthcare Life Sciences). The dissociation constant was determined by using the software of GraphPad Prism.

Cross linking of $Taz1_D$

The protein of $Taz1_D$ was incubated with various concentration of DSS cross-linker (Disuccinimidyl suberate, Thermo Scientific) at 25 °C in PBS. After incubation for 30 min, 10 μl aliquots of the reaction mixture were withdrawn, diluted with 10 μl of water and 5 μl of SDS loading dye, and run on 15% SDS-PAGE visualized with Coomassie brilliant blue stain.

Crystallization, data collection and structure determination

The $Taz1_{TRFH}$ protein was crystallized by hanging-drop-vapor-diffusion at 4 °C. The precipitant/well solution contained 100 mM Tris-HCl pH7.0, 18% PEG 3350, 300mM

KNO_3, 10 mM $Co(NH_4)_6Cl_3$ and 10 mM DTT. Crystals were gradually transferred to a harvesting solution containing 100 mM Tris-HCl pH7.0, 25% PEG 3350, 300mM KNO_3, 10 mM $Co(NH_4)_6Cl_3$ and 25% glycerol before flash-frozen in liquid nitrogen for storage and data collection under cryogenic conditions (100 K). Se-Met-MAD datasets with the resolution of 2.7 Å were collected at beam line 21ID-D at APS and processed using HKL2000 (Otwinowski and Minor, 1997). Crystals belong to space group *C*222 and contain one $Taz1_{TRFH}$ per asymmetric unit. Two selenium sites were located and refined, and MAD phases calculated using SHARP (E. d. La Fortelle, G. Bricogne, 1997). A model was automatically built into the modified experimental electron density using ARP/WARP (V. S. Lamzin, A. Perrakis, K. S. Wilson, 2001); the model was then further refined in CNS [25] with manual rebuilding using program O [26].

The $Taz1_D$ protein was crystallized by sitting-drop-vapor-diffusion at 4 °C. the precipitant/well solution contained 2.3 M $(NH_4)_2SO4$, 100 mM Bicine pH 8.5 and 10 mM DTT. Crystals were gradually transferred to a harvesting solution. Final concentration of 125 mM NaI was added into harvesting solution 1 minute before the flash-frozen in liquid nitrogen for storage and data collection under cryogenic conditions (100 K). Crystals belong to space group $P4_32_12$ with one $Taz1_D$ monomer per asymmetric unit. Datasets with the resolution of 1.5 Å were collected at beam line 21ID-D at APS and processed using HKL2000 (Otwinowski and Minor, 1997). Native-/I- datasets were combined and scaled in ccp4i and then were input into SHARP (E. d. La Fortelle, G. Bricogne, 1997). The initial SIR map was significantly improved by solvent flattening. A model was automatically built into the modified experimental electron density using ARP/WARP (V. S. Lamzin, A. Perrakis, K. S. Wilson, 2001); the model was then further

refined using simulated-annealing and positional refinement in CNS [25] with manual rebuilding using program O [26].

Yeast two-hybrid assay

The yeast two-hybrid assays were performed using L40 strains harboring pBTM116 and pACT2 (Clonetech) fusion plasmids. The colonies containing both plasmids were selected on –Leu –Trp plates. The β-galactosidase activities were measured according to Clontech MATCHMAKER library protocol.

NMR Spectroscopy

The NMR experiments were carried out at 25 ºC on Bruker 600- and 800-MHz spectrometers equipped with four RF channels and triple resonance pulsed-field gradient cryoprobes. The chemical shifts were referenced to internal 2, 2-dimethyl-2-silapentanesulfonic acid (DSS). The samples were prepared with 1.5 mM protein dissolved in a buffer of 90%H_2O/10% D_2O containing 20 mM sodium phosphate (pH 6.5) and 50 mM NaCl. All NMR spectra were processed with NMRPipe [27] and analyzed with Sparky (T.D. Goddard and D.G. Kneller, SPARKY 3, University of California, San Francisco). Two-dimensional ^{15}N- and ^{13}C-edited HSQC, (HB)CB(CGCD)HD, (HB)CB(CGCDCE)HE, and three-dimensional HNCACB, CBCA(CO)NH, HNCO, HN(CA)CO, C(CCO)NH, H(CCCO)NH, HCCH-TOCSY and CCH-COSY spectra were recorded to obtain the chemical shift assignments of backbone and sidechain atoms. The three-dimensional ^{15}N- and ^{13}C-edited NOESY-HSQC spectra

(mixing time 100 ms) were collected to generate the distance restraints for structure calculations.

Structure Calculation

Initial structures were calculated using the program ARIA 2.2 [28], NOE peaks were assigned with SANE [29] and CYANA 2.1 [30], and the final structures were refined with Amber 9.0 (D.A. Case et al. 2006). Distance restraints were derived from interproton NOEs. Backbone dihedral angle (φ and ψ) restraints were generated from chemical shift data using TALOS [31]. Hydrogen bond restraints were determined using the secondary structure information from CSI [32] and confirmed by intermediate range NOEs. The 20 lowest energy structures from ARIA were selected as models for SANE to extend the NOE assignments. The final set of distance restraints were obtained after several rounds of SANE/CYANA calculations. Two hundred structures from CYANA were refined by restrained molecular dynamics calculations with Amber using generalized Born salvation model to account for solvent effects. The 20 refined structures with the lowest energy were analyzed using [33] and the averages from three individual transformants were reported.

Yeast Strains

A region of *taz1*$^{+}$ open reading frame (from bases pair 315 to 501) has been replaced by *ura4*$^{+}$ gene to generate a *taz1::ura* (Tomita K., unpublished data). Vectors *pACT2-taz1-*$L_{431}R$ / *taz1-*$V_{434}W$ / *taz1-*$L_{438}W$ / *taz1-*$L_{445}R$ as well as vectors *pET-SUMO-Taz1* $I_{379}R$ / *Taz1-*$L_{380}R$ / *Taz1-*$L_{383}R$ / *Taz1-*$L_{387}R$ were digested with BamHI / XhoI, generating a

taz1 full length fragment containing a set of point mutation in Taz1 dimerization domain (*taz1-L_{431}R / taz1-V_{434}W / taz1-L_{438}W / taz1-L_{445}R*) and Taz1-Rap1 interaction domain (*taz1 I_{379}R / taz1-L_{380}R / taz1-L_{383}R / taz1-L_{387}R*). *taz1::ura* strain was transformed with the BamHI / XhoI fragments from *pACT2* and *pET-SUMO* and plated on replica-plated twice on EMM + FOA to select for the loss of *ura4+* gene. Transformants were screened by PCR using two sets of primers. Primers 59/60 hybridize outside *taz1+* ORF; primers 146/60 hybridize inside and outside *taz1+* ORF respectively.

Measurement of telomere length

Telomere length was measured by Southern hybridization according to the procedure described by Cooper J.P. *et al* 1997 [34]. Briefly, genomic *S.pombe* DNA was prepared by glass beads lysis method, digested with EcoRI, resolved on 1% agarose gel, transferred to Hybond membrane (Amersham) and hybridized at 65°C to ^{32}P random primed telomeric probe (SacI-PstI fragment from pIRT2-TELO vector).

Figure 3.1 Characterization of $Taz1_{TRFH}$

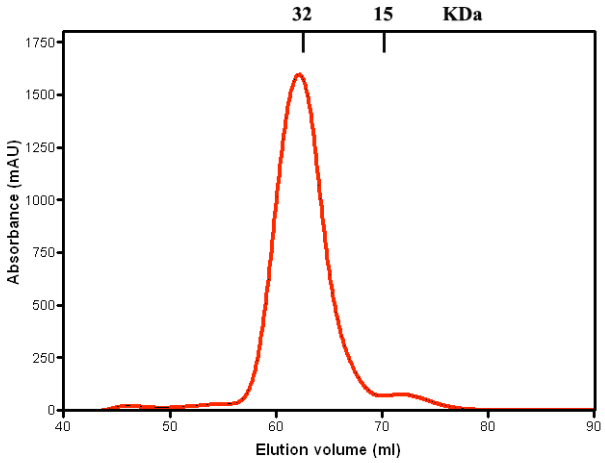
(a) $Taz1_{117-391}$ is a monomor in solution. Gel filtration chromatography profile (Hiload Superdex 75) of the $Taz1_{117-391}$ protein. Elution positions of 15 and 32 KDa protein markers are indicated.

(b) Limited proteolysis of $Taz1_{113-400}$ by Trypsin, Elastase, Subtilisin and Glu-C. Reaction time and different proteases as labeled.

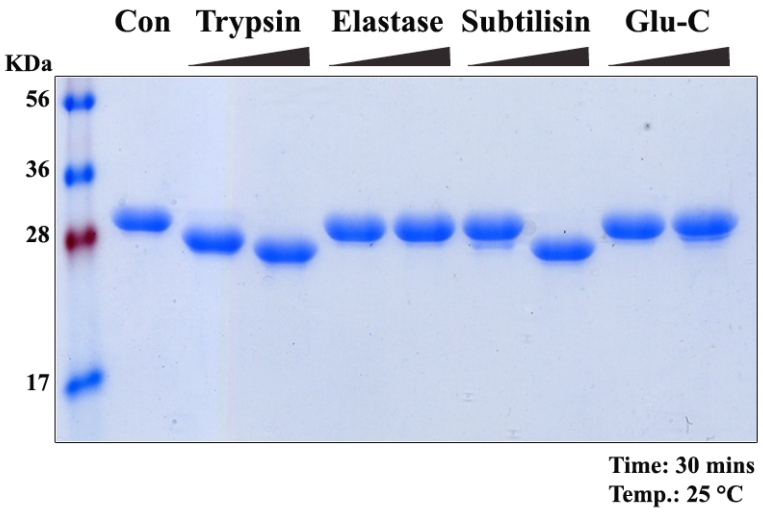
(c) SDS-PAGE time course of Trypsin cleavage of $Taz1_{113-400}$. Lanes in minutes of time of the reaction as labeled.

(d) The molecular weight of the stable fragment resulted from **(c)** was identified by mass spectrometry, which is 29360.2 $\pm$ 30 Da.

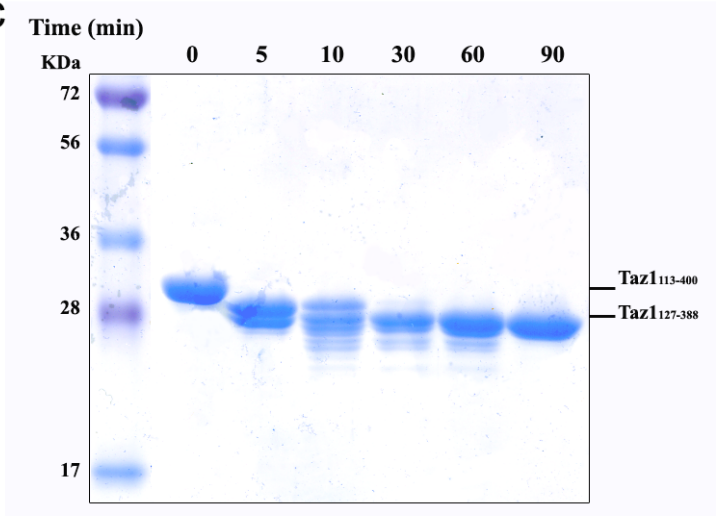
a


32
15
KDa
1750
1500
1250
1000
750
500
250
0
Absorbance (mAU)
40
50
60
70
80
90
Elution volume (ml)

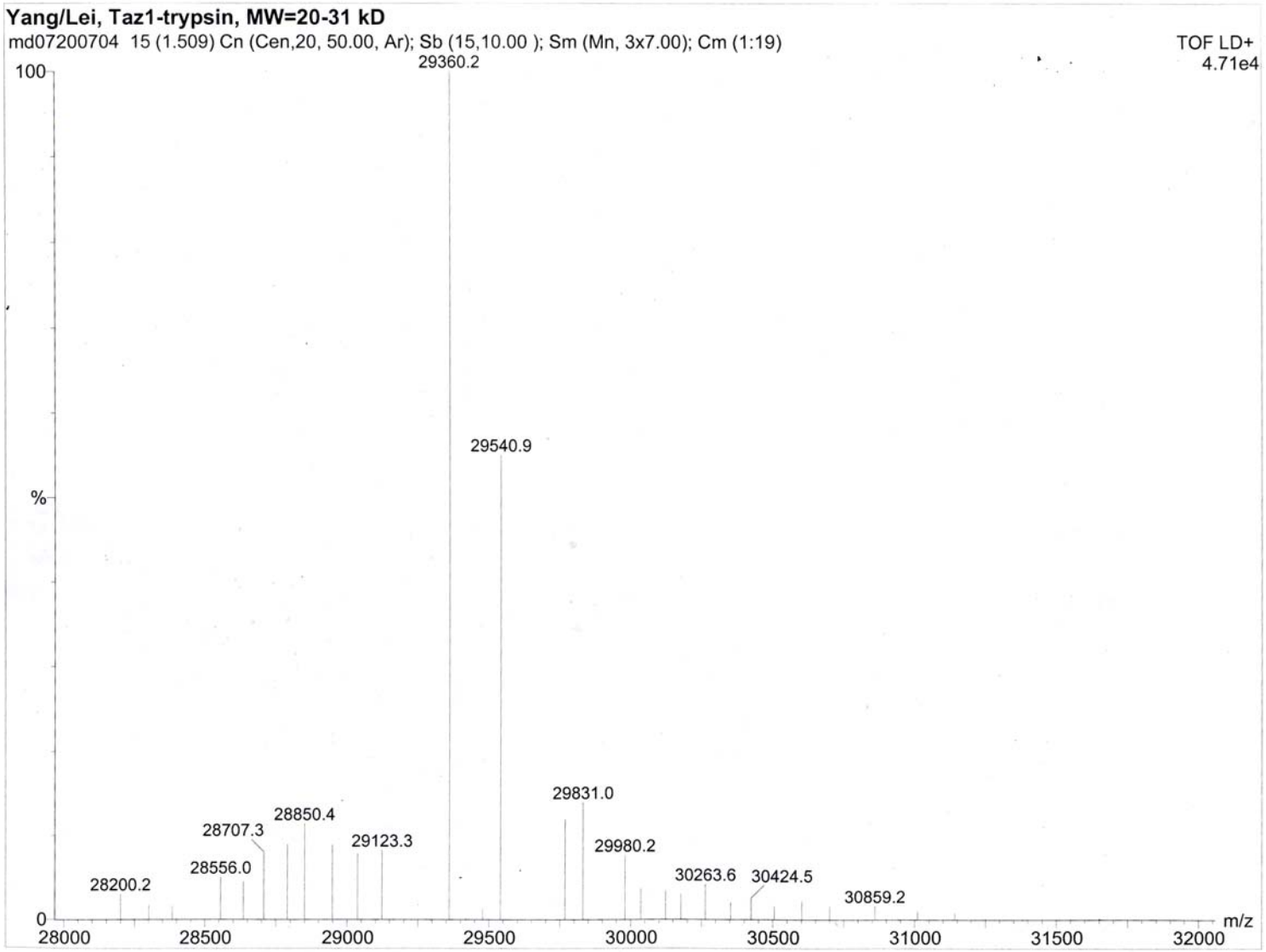
b


Con
Trypsin
Elastase
Subtilisin
Glu-C
KDa
56
36
28
17
Time: 30 mins
Temp.: 25 °C

c


Time (min)
KDa
0
5
10
30
60
90
72
56
36
28
17
Taz1113-400
Taz1127-388

d


Yang/Lei, Taz1-trypsin, MW=20-31 kD
md07200704 15 (1.509) Cn (Cen,20, 50.00, Ar); Sb (15,10.00); Sm (Mn, 3x7.00); Cm (1:19)
TOF LD+
4.71e4
100
%
0
29360.2
29540.9
29831.0
28850.4
28707.3
29123.3
29980.2
28556.0
28200.2
30263.6
30424.5
30859.2
m/z
28000
28500
29000
29500
30000
30500
31000
31500
32000

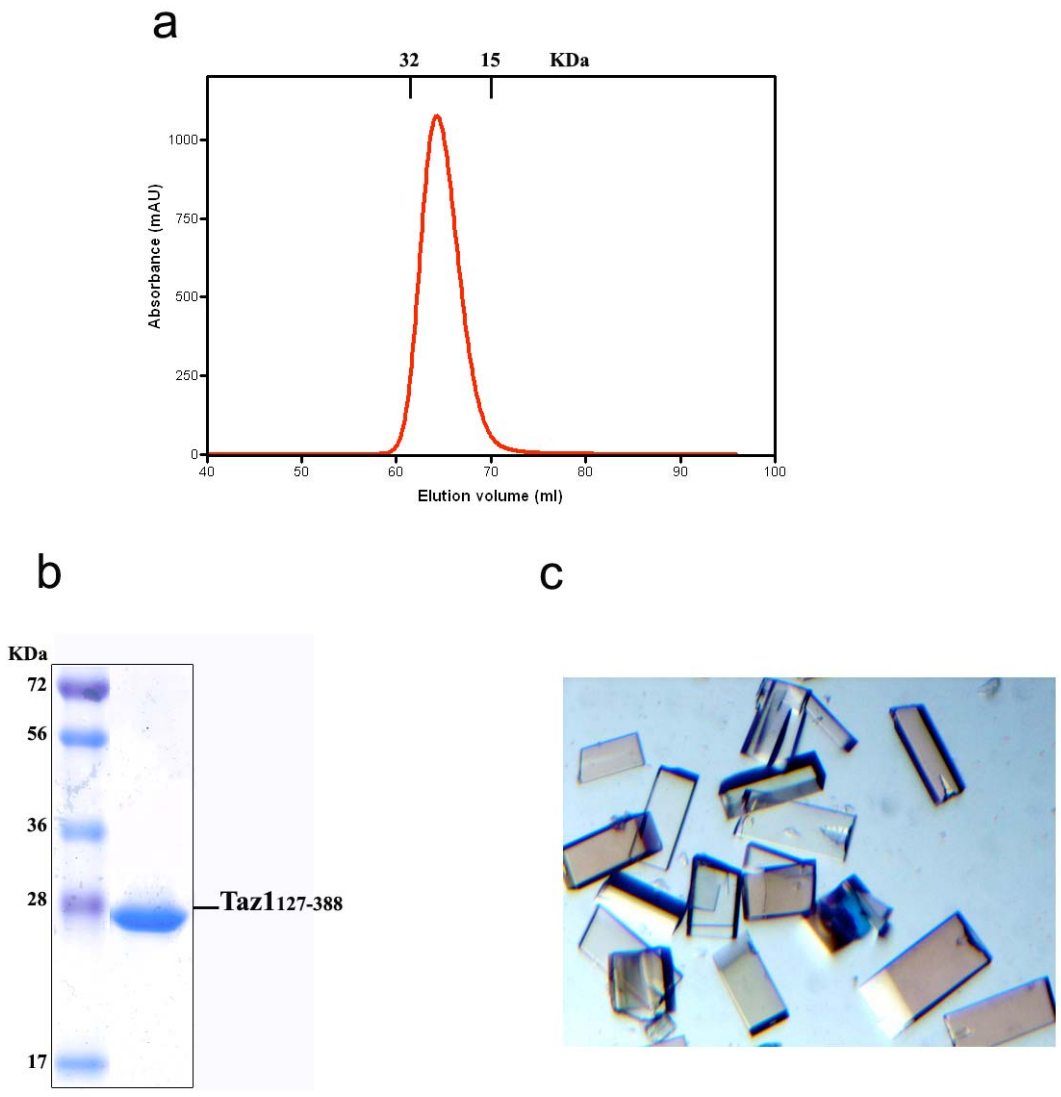


Figure 3.2 Purification and crystallization of Taz1$_{TRFH}$
(a) Gel filtration chromatography profile (Hiload Superdex 75) of the Taz1$_{127-388}$ protein. Elution positions of 15 and 32 KDa protein markers are indicated.
(b) SDS-PAGE of purified protein of Taz1$_{127-388}$ for crystallization.
(c) Crystals of Taz1$_{127-388}$.

Table 3.1 Data collection, phasing and refinement statistics for $Taz1_{TRFH}$

	$Taz1_{TRFH}$			
Data collection				
	Se Peak	**Se Inflection**		**Native**
Space group		*C*222		*C*222
Cell dimensions				
a, *b*, *c* (Å)	68.006, 160.354, 49.941			68.008, 160.097, 50.080
α, β, γ (º)	90, 90, 90			90, 90, 90
Wavelength (Å)	0.97935	0.97951		0.97935
Resolution (Å)	50-2.7	50-2.7		100-2.25
R_{merge} (%) (high res. shell)	0.063 (0.518)	0.055 (0.368)		0.043 (0.373)
I/σ (high res. cell)	52.5 (4.9)	55.7 (8.6)		57.6 (3.6)
Completeness (%) (high res. shell)	99.4 (96.8)	99.6 (98.8)		96.6 (80.3)
Redundancy (high res. shell)	16.4 (12.9)	16.8 (15.6)		6.7 (5.4)
Phasing				
	Anomalous	Anomalous	Isomorphous	
Acentric Phasing Power	2.051	1.471	0.625	
Centric Phasing Power			0.608	
Acentric reflections FOM	0.304			
Centric reflections FOM	0.117			
Refinement				
Resolution (Å)				50-2.3
No. reflections				12401
R_{work}/ R_{free} (%)				23.66/28.57
B-factors (Å^2)				
Protein				77.6
Water				66.9
R.m.s deviations				
Bond lengths (Å)				0.006820
Bond angles (º)				1.28085

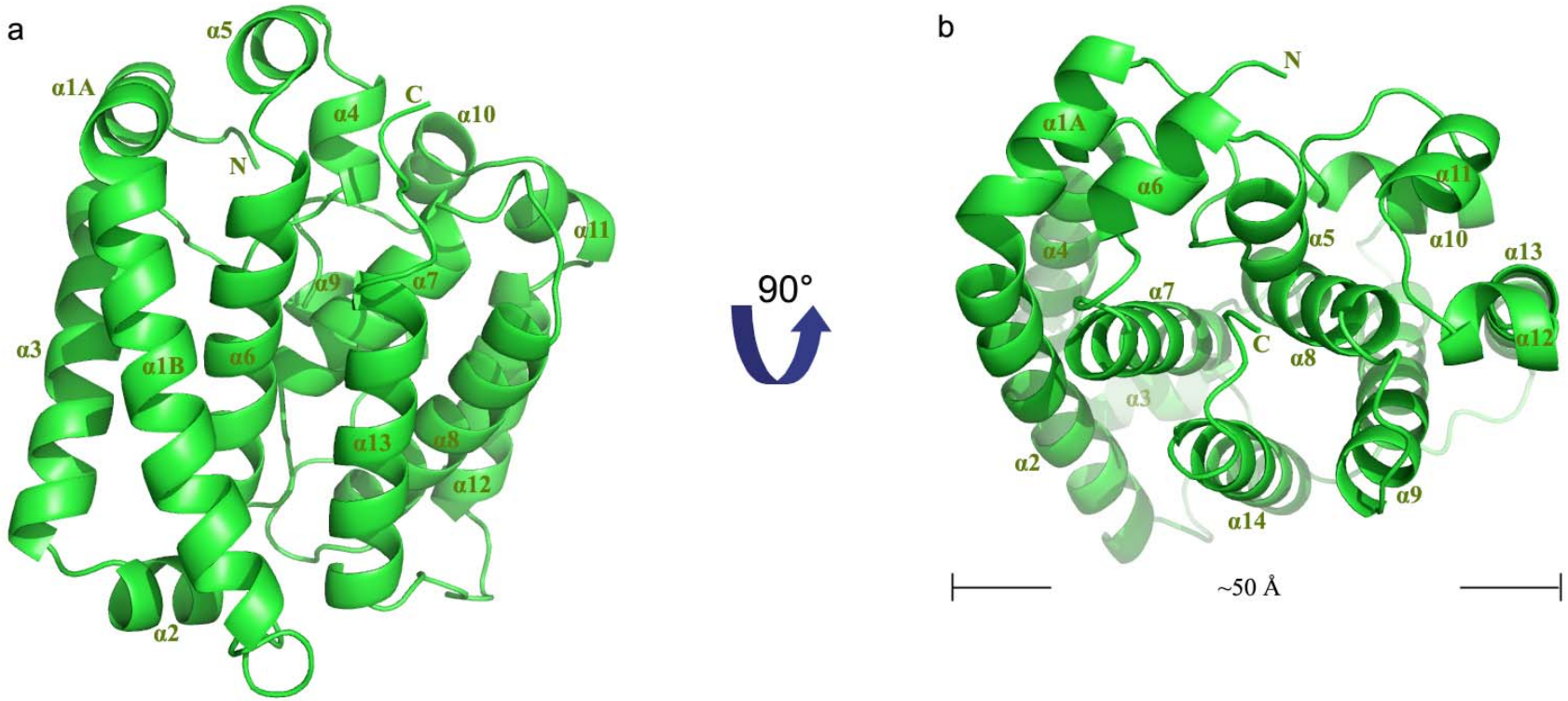
Figure 3.3 Crystal structure of $Taz1_{TRFH}$

(a), (b) Overall structure of $Taz1_{TRFH}$. $Taz1_{TRFH}$ consists 14 α helices: α1B, α3, α6, α7, α8, α9, α12, and α13 form an eight-helix bundle, while the other six short helices α1A, α2, α4, α5, α10, and α11, cap the exposed hydrophobic surface from the sides of the central helical bundle. The helices are organized so that the entire domain resembles the shape of a slightly curved hockey puck with a diameter of ~ 50 Å.

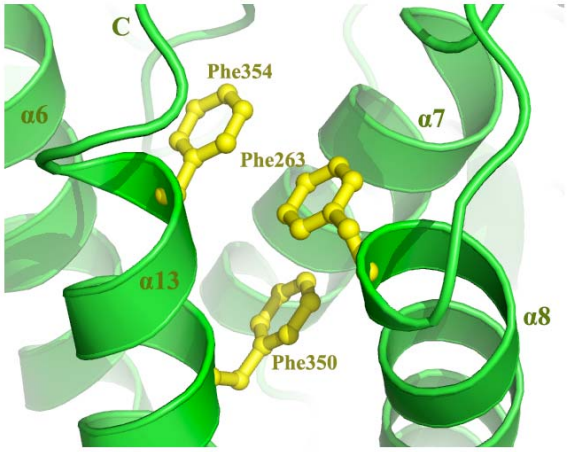
(c) Three phenylalanines (Phe263 from α8, Phe350 and Phe354 from α13) stack with other and keep the interaction between helix α8 and α13.

(d) $Taz1_{TRFH}$ is in surface representation and colored according to its electrostatic potential (positive potential, blue; negative potential, red). $Taz1_{TRFH}$ surface is highly negative charged.

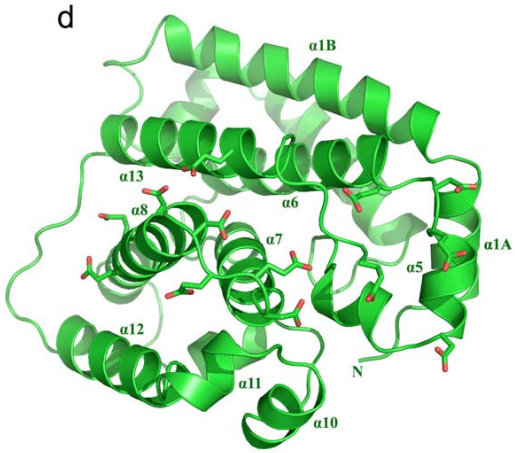
(e) On the surface of $Taz1_{TRFH}$, there is a cluster of 13 acidic residues that are distributed in several peripheral short helices.


a
α5
α1A
C
α4
α10
N
α11
α9
α7
α3
α1B
α6
α13
α8
α12
α2
90°
b
N
α1A
α6
α11
α4
α5
α10
α13
α7
C
α8
α12
α3
α2
α9
α14
~50 Å

c


C
Phe354
α6
α7
Phe263
α13
α8
Phe350

d


α1B
α13
α6
α8
α7
α1A
α5
α12
N
α11
α10

e

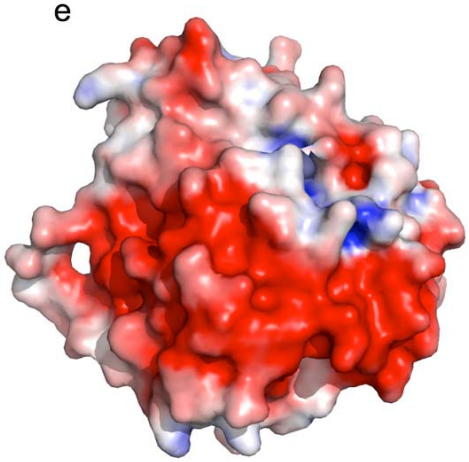
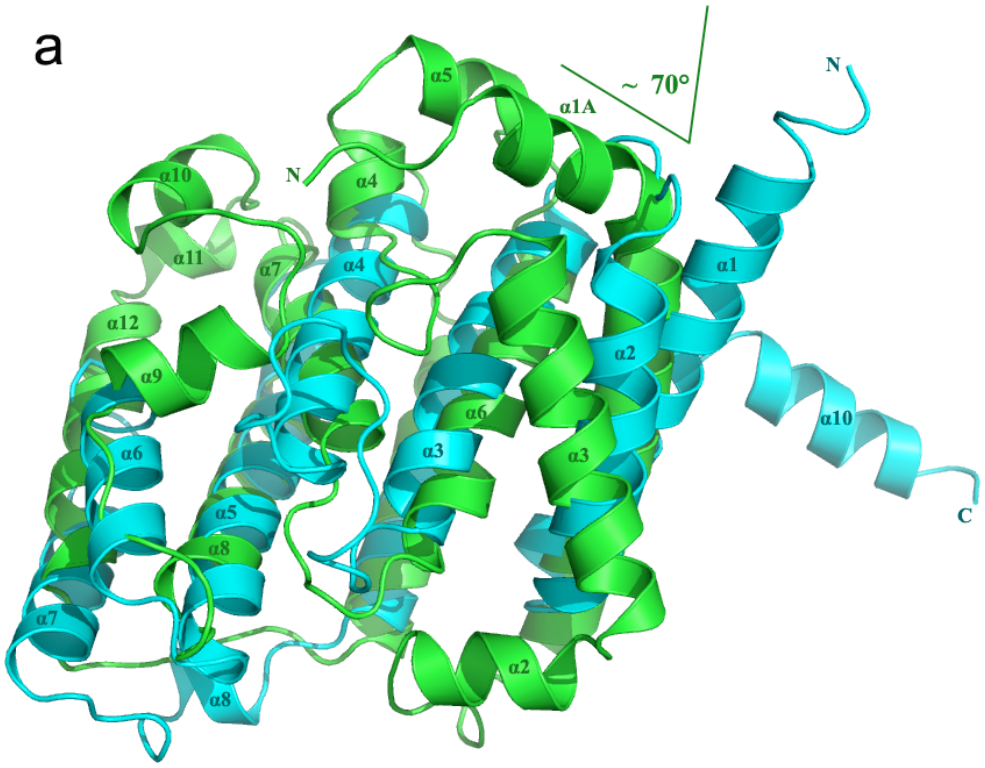
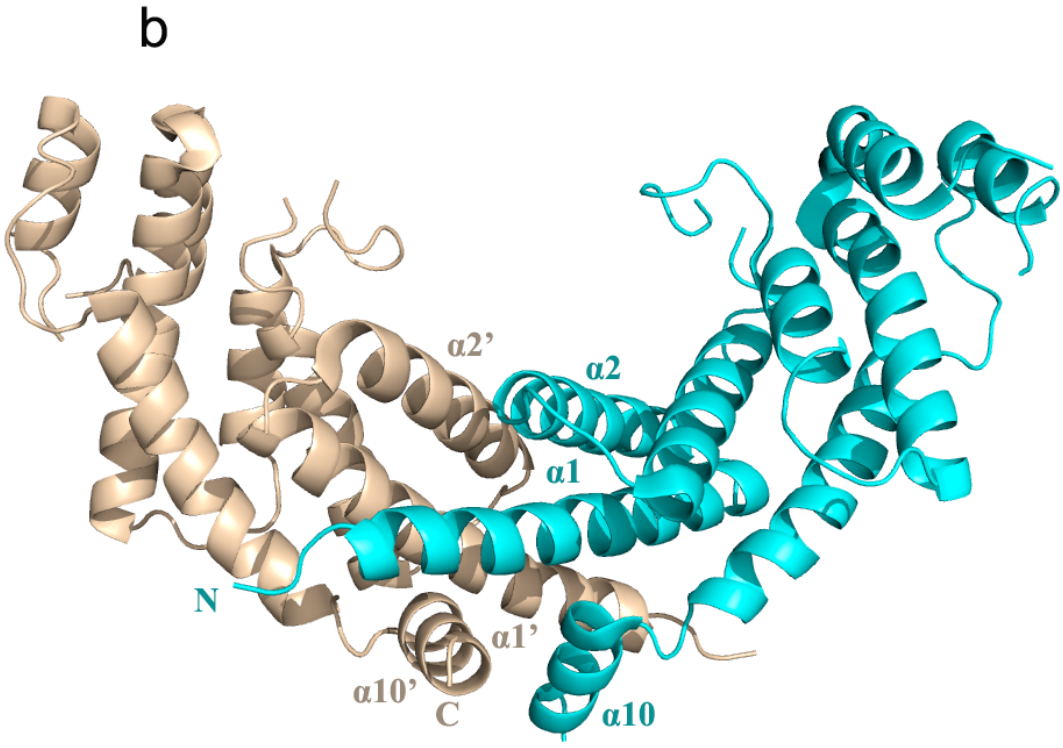


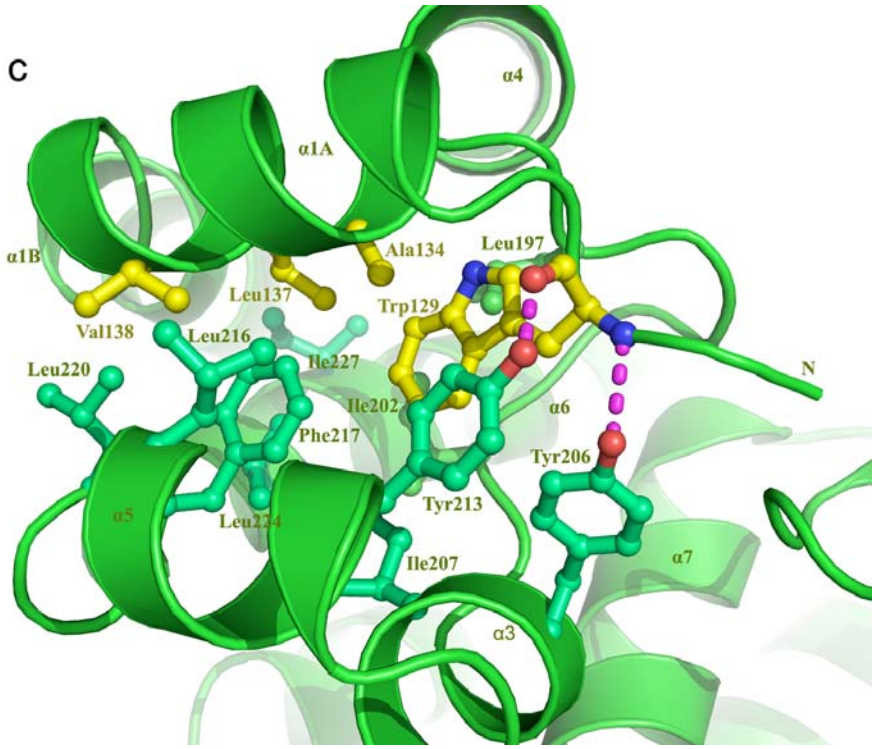
Figure 3.4 Comparison of $Taz1_{TRFH}$ with $TRF1/TRF2_{TRFH}$
(a) Superposition of the $Taz1^{TRFH}$ structure on the crystal structure of TRF2 [35]. Taz1 is colored in green and $TRF2^{TRFH}$ is colored in cyan. The α1 helix of $Taz1_{TRFH}$ is severely bent (~ 70°).
(b) The dimerization of $TRF1_{TRFH}$. The N-terminal portion of α1 extends outside of the helical core and interacts with helix α10 from the other molecule in the dimer.
(c) The bended α1A helix and the N-terminal tail of $Taz1_{TRFH}$ fits into a hydrophobic groove formed by helices α3, α4, and α5. Two hydrogen bonding interactions are formed between main-chain atoms of Trp129 and the hydroxyl groups of Tyr206 and Tyr213, respectively.
(d) Secondary structure assignments from the crystal structure of $Taz1_{TRFH}$ and $TRF1/TRF2_{TRFH}$ are shown as colored ($Taz1_{TRFH}$ is green, $TRF1/TRF2_{TRFH}$ are blue) cylinders (α helices,) above the aligned sequences. Highlighted areas indicate the conserved residues in the TRFH domain of Taz1 and TRF1/TRF2. The dashed lines show the disordered regions.



Taz1TRFH / TRF2TRFH



TRF1TRFH dimer


C
α4
α1A
α1B
Ala134
Leu197
Leu137
Trp129
Val138
Leu216
Ile227
Leu220
N
Ile202
α6
Phe217
Tyr206
Tyr213
α5
Leu234
α7
Ile207
α3

d

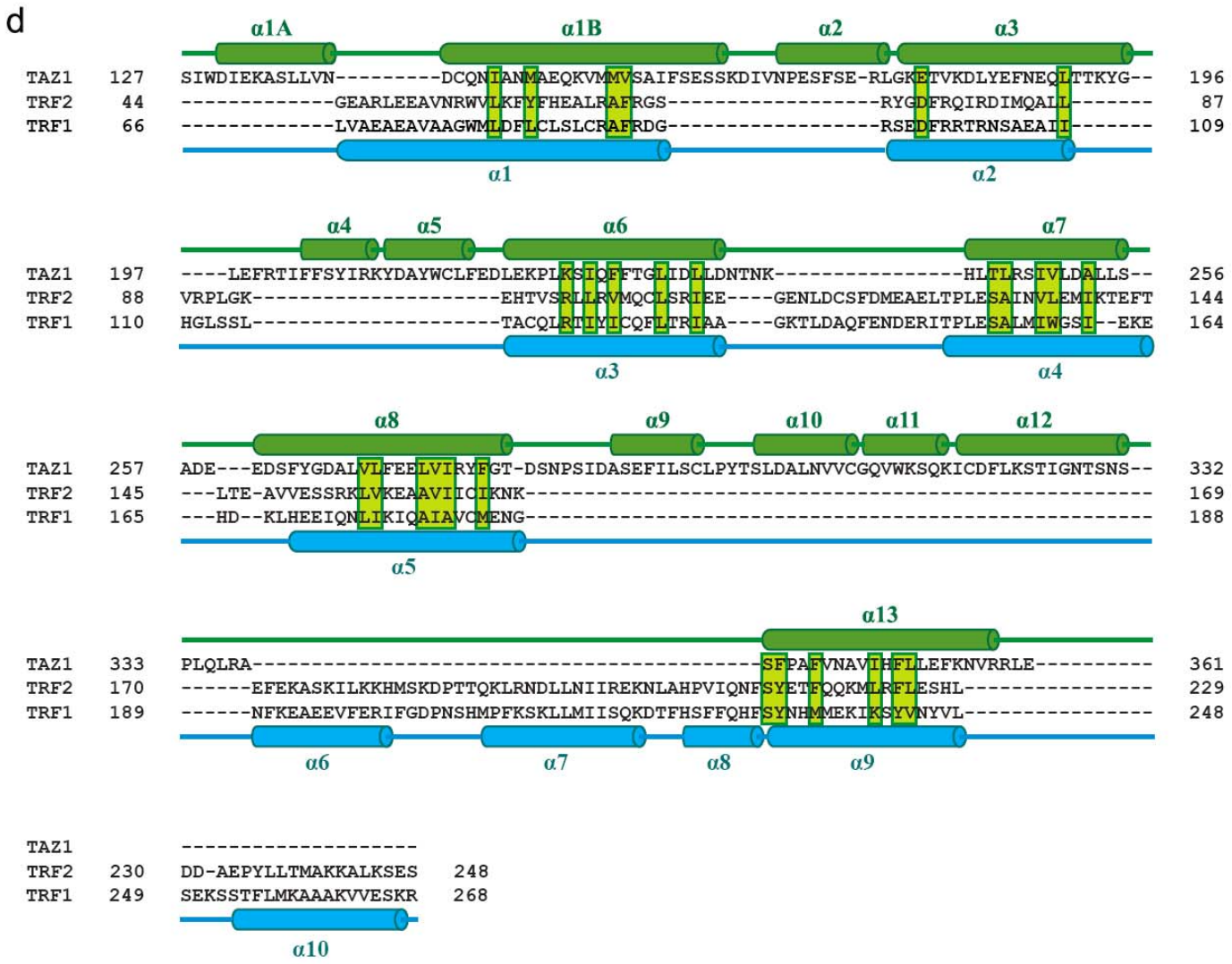
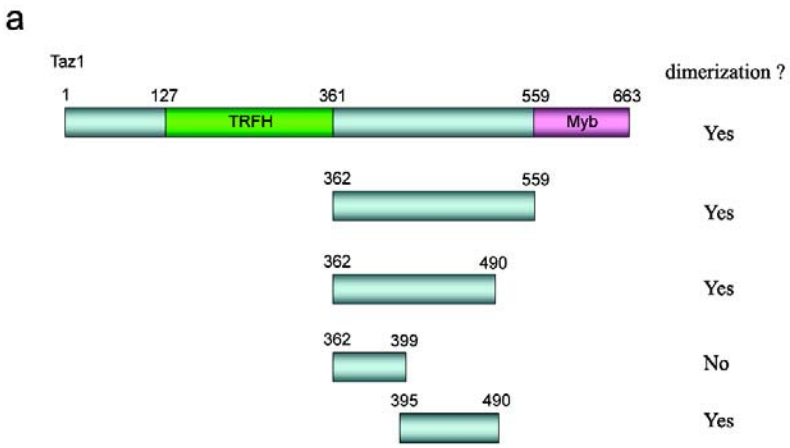
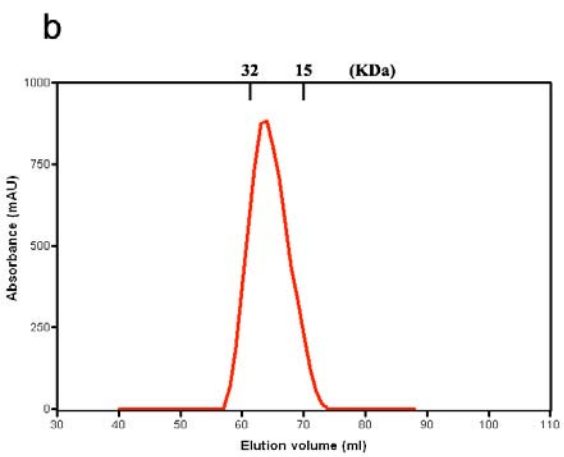
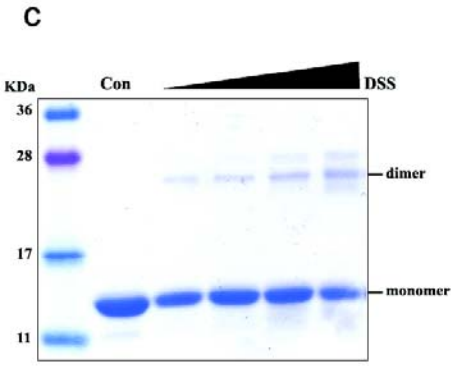

α1A
α1B
α2
α3
TAZ1 127 SIWDIEKASLLVN---------DCQNIANMAEQKVMMVSAIFSESSKDIVNPESFSE-RLGKETVKDLYEFNEQLTTKYG-- 196
TRF2 44 -------------GEARLEEAVNRWVLKFYFHEALRAFRGS------------------RYGDFRQIRDIMQALL------- 87
TRF1 66 -------------LVAEAEAVAAGWMLDFLCLSLCRAFRDG------------------RSEDFRRTRNSAEAII------- 109
α1
α2
α4
α5
α6
α7
TAZ1 197 ----LEFRTIFFSYIRKYDAYWCLFEDLEKPLKSIQFFTGLIDLLDNTNK----------------HLTLRSIVLDALLS-- 256
TRF2 88 VRPLGK----------------------EHTVSRLLRVMQCLSRIEE----GENLDCSFDMEAELTPLESAINVLEMIKTEFT 144
TRF1 110 HGLSSL----------------------TACQLRTIYICQFLTRIAA----GKTLDAQFENDERITPLESALMIWGSL--EKE 164
α3
α4
α8
α9
α10
α11
α12
TAZ1 257 ADE---EDSFYGDALVLFEELVIRYFGT-DSNPSIDASEFILSCLPYTSLDALNVVCGQVWKSQKICDFLKSTIGNTSNS-- 332
TRF2 145 ---LTE-AVVESSRKLVKEAAVIICIKNK-- 169
TRF1 165 ---HD--KLHEEIQNLIKIQAIAVCMENG-- 188
α5
α13
TAZ1 333 PLQLRA---SFPAFVNAVIHFLLEFKNVRRLE---------- 361
TRF2 170 ------EFEKASKILKKHMSKDPTTQKLRNDLLNIIREKNLAHPVIQNFSYETFQQKMLRFLESHL---------- 229
TRF1 189 ------NFKEAEEVFERIFGDPNSHMPFKSKLLMIISQKDTFHSFFQHFSYNHMMEKIKSYVNYVL---------- 248
α6
α7
α8
α9
TAZ1 ----------
TRF2 230 DD-AEPYLLTMAKKALKSES 248
TRF1 249 SEKSSTFLMKAAAKVVESKR 268
α10

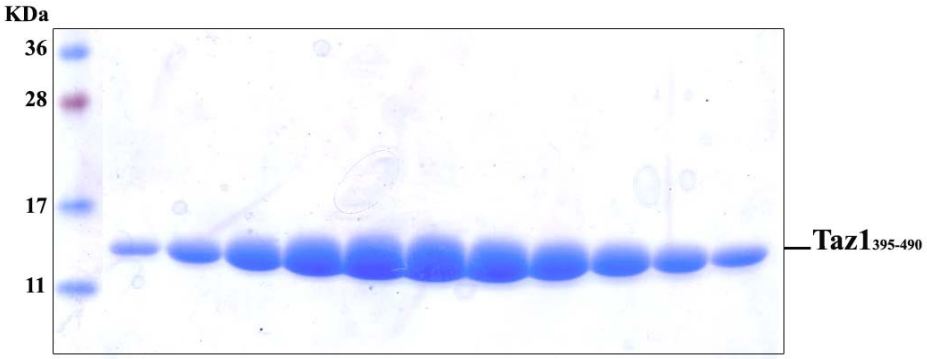
Figure 3.5 Characterization of Taz1 dimerization domain
(a) Various constructs are designed for determination of Taz1D. Dimerization is tested via gel filtration chromatography and chemical cross-linking analysis.
(b) Gel filtration chromatography profile (Hiload Superdex 75) of the $Taz1_{395-490}$ protein. Elution positions of 15 and 32 KDa protein markers are indicated.
(c) Dimerization of $Taz1_{395-490}$ protein that cross-linked by DSS cross-link reagent (Disuccinimidyl suberate, Thermo Scientific). With increased amount of DSS added, more $Taz1_{395-490}$ protein is cross-linked as the dimer.
(d) SDS-PAGE of $Taz1_{395-490}$ protein fractions from **(b)**.
(e) Crystals of $Taz1_{395-490}$.


a
Taz1
dimerization ?
1
127
361
559
663
TRFH
Myb
Yes
362
559
Yes
362
490
Yes
362
399
No
395
490
Yes


b
32
15
(KDa)
1000
750
500
250
0
Absorbance (mAU)
30
40
50
60
70
80
90
100
110
Elution volume (ml)


c
KDa
Con
DSS
36
28
17
11
dimer
monomer

d



e

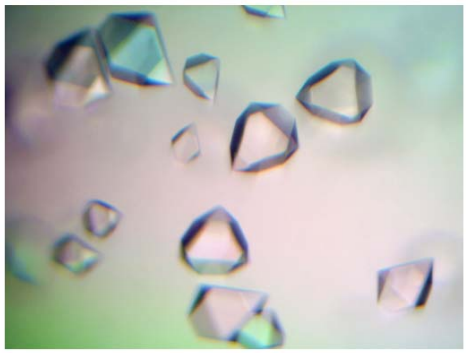
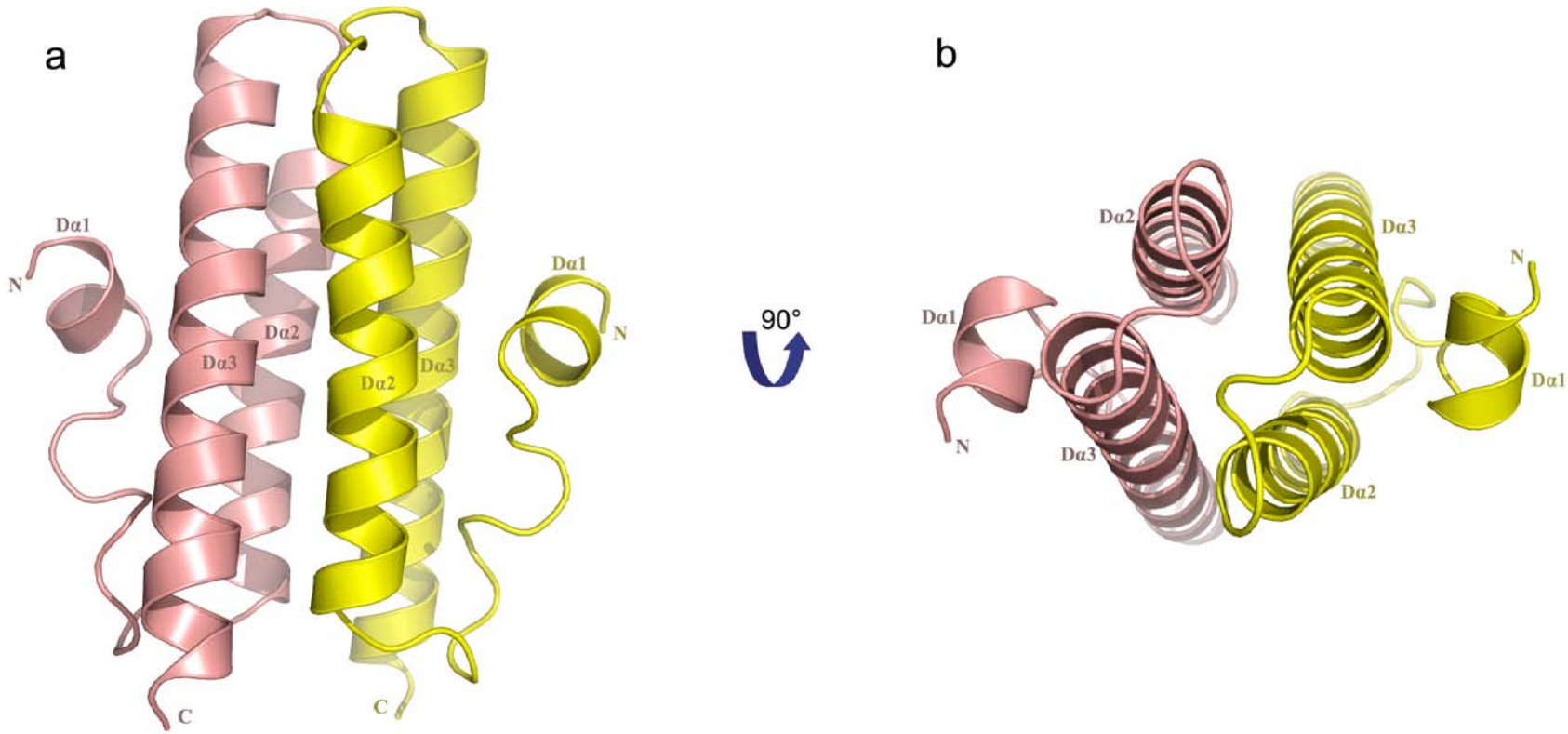
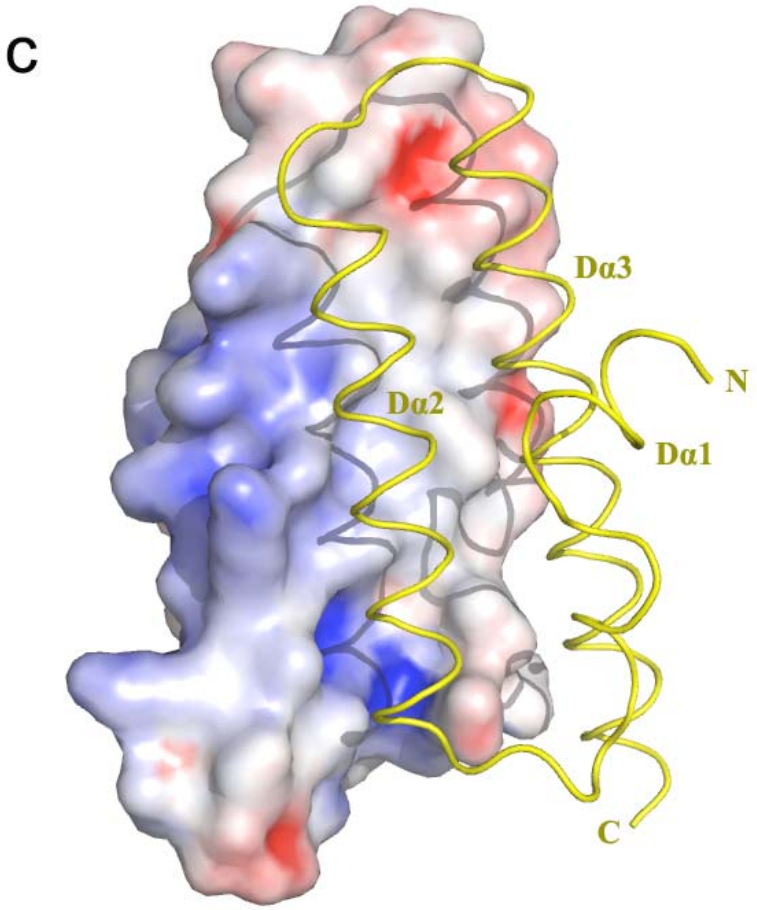


Table 3.2 Data collection, phasing and refinement statistics for $Taz1_D$

	$Taz1_D$	
Data collection		
	NaI	**Native**
Space group	$P4_32_12$	$P4_32_12$
Cell dimensions		
a, *b*, *c* (Å)	45.170, 45.170, 79.341	45.078, 45.078, 78.687
α, β, γ (º)	90, 90, 90	90, 90, 90
Wavelength (Å)	0.97872	0.97872
Resolution (Å)	100-1.3	100-1.5
R_{merge} (%) (high res. shell)	0.086 (0.355)	0.046 (0.271)
I/σ (high res. cell)	54.7 (2.63)	38.6 (2.1)
Completeness (%) (high res. shell)	95.3 (64.7)	97.1 (85.9)
Redundancy (high res. shell)	11.6 (4.1)	6.5 (4.0)
Phasing		
Acentric Phasing Power	0.623	
Centric Phasing Power	0.634	
Acentric reflections FOM	0.129	
Centric reflections FOM	0.185	
Refinement		
Resolution (Å)		50-1.5
No. reflections		12913
R_{work}/ R_{free} (%)		25.0/26.7
B-factors (Å^2)		
Protein		24.4
Water		40.9
R.m.s deviations		
Bond lengths (Å)		0.003554
Bond angles (º)		1.02094

Figure 3.6 Crystal structure of Taz1 dimerizatin domain
(a)-(c) Taz1 dimerization domain is a four-helix bundle structure, resulting the burial of a total of 2514.96 $Å^2$ of the solvent accessible surface area. In (a) and (b), one monomer is colored yellow and the other monomer is colored light red. In (c), one monomer is in surface representation and colored according to its electrostatic potential (positive potential, blue; negative potential, red).
(d) The side chains of L415, L417, I420, and I425 insert into a hydrophobic groove formed by residues from helices Dα2 and Dα3. L415 has two hydrogen-bonding interactions with Q456 and K440.


a
Dα1
N
Dα3
Dα2
C
Dα2
Dα3
Dα1
N
C
90°
b
Dα2
Dα3
N
Dα1
Dα1
N
Dα3
Dα2


c
Dα3
N
Dα2
Dα1
C

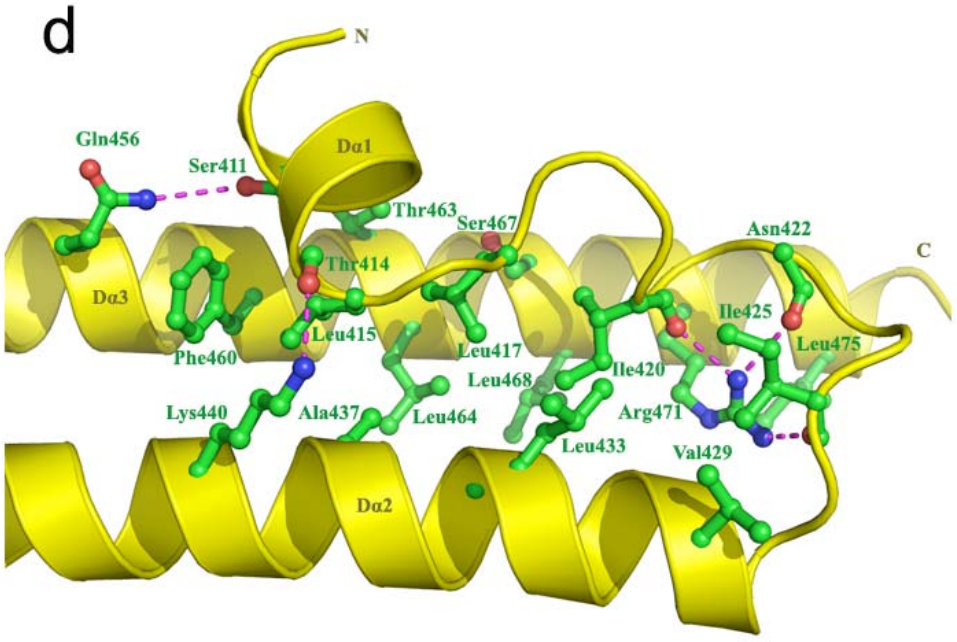

d
N
Gln456
Ser411
Dα1
Thr463
Ser467
Asn422
C
Thr414
Dα3
Ile425
Leu475
Leu415
Leu417
Phe460
Leu468
Ile420
Lys440
Ala437
Leu464
Arg471
Leu433
Val429
Dα2

Figure 3.7 Dimerization interface of $Taz1_D$
In **(a, b)**, the hydrophobic packing contact at the dimeric interface of $Taz1_D$ is extensive, with 12 residues from one molecule interacting with those from the other molecule. Besides, intermolecular electrostatic interactions and three hydrogen-bonding interactions provide additional specificity and stability to the structure. In **(a)**, one monomer is in surface representation and colored according to its electrostatic potential (positive potential, blue; negative potential, red).

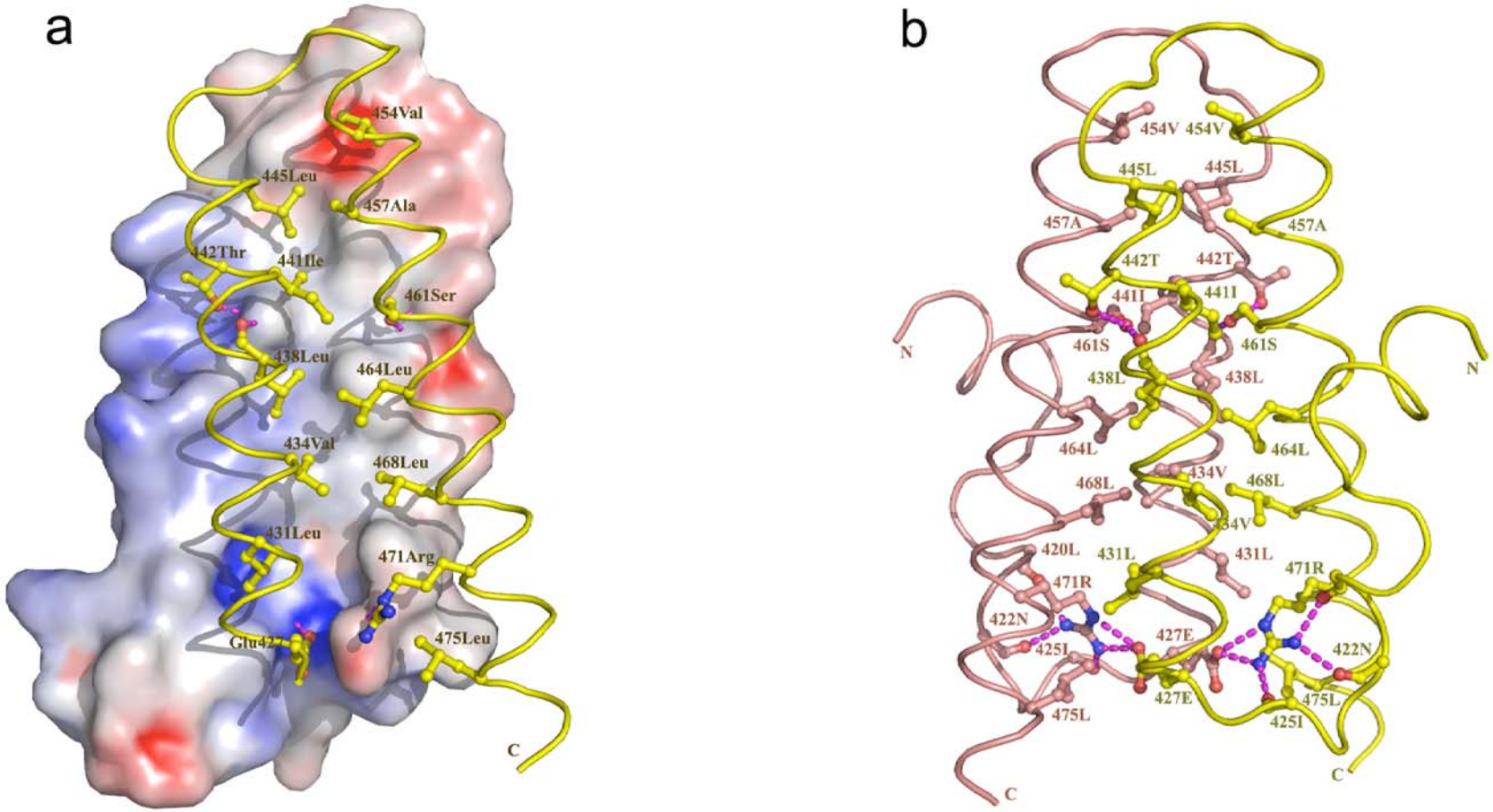
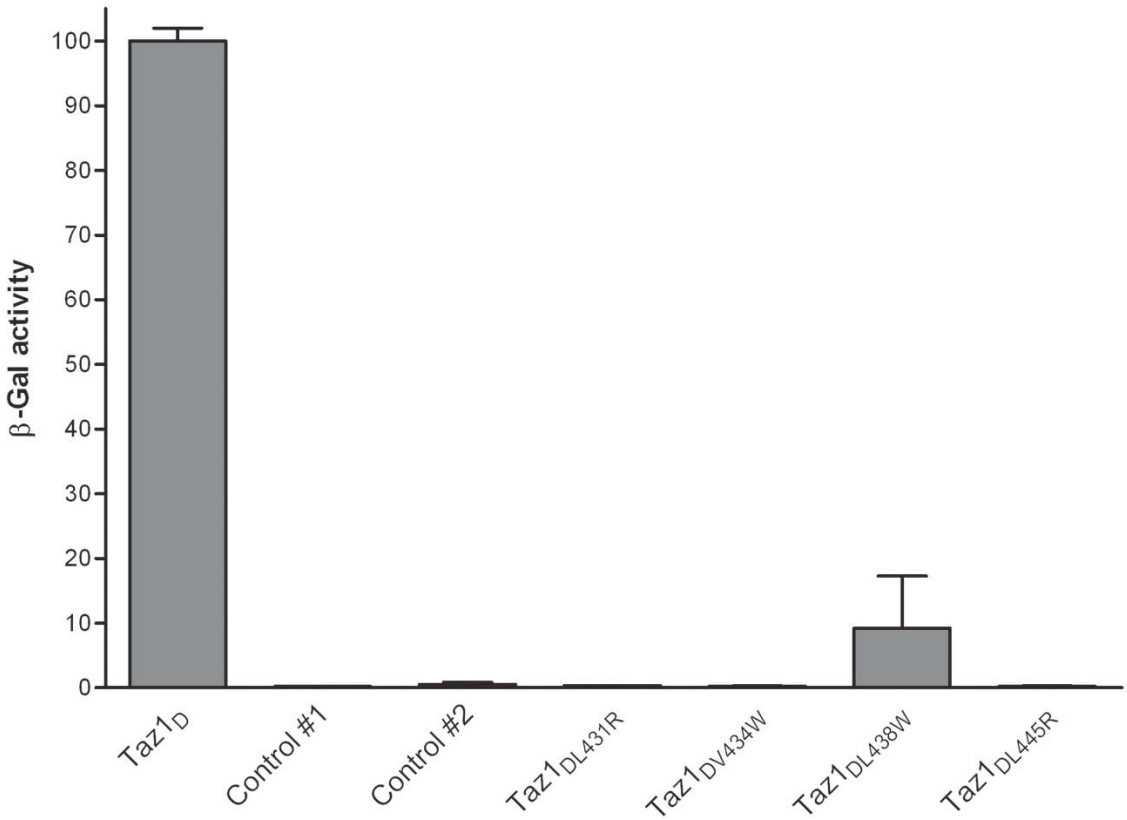

a
454Val
445Leu
457Ala
442Thr
441Ile
461Ser
438Leu
464Leu
434Val
468Leu
431Leu
471Arg
Glu427
475Leu
C
b
454V 454V
445L 445L
457A
457A
442T 442T
441I 441I
461S
461S
N
N
438L 438L
464L
464L
468L
434V
468L
434V
420L
431L
431L
471R
471R
422N
427E
422N
425I
427E
475L
475L
425I
C
C

Figure 3.8 Effects of the mutations on the Taz1 dimerization interface
(a) In yeast two-hybrid, interaction of LexA-$Taz1_D$ with GAD- $Taz1_D$ was measured as β-galactosidase activity. Data are averaged of three independent β-galactosidase measurements.
(b) Yeast two-hybrid results consistent with the observation that the binding pocket for Leu438 in the crystal structure is deeper than that for Val434. One monomer is in surface representation and colored according to its electrostatic potential (positive potential, blue; negative potential, red).

a


β-Gal activity
100
90
80
70
60
50
40
30
20
10
0
Taz1D
Control #1
Control #2
Taz1DL431R
Taz1DV434W
Taz1DL438W
Taz1DL445R

b

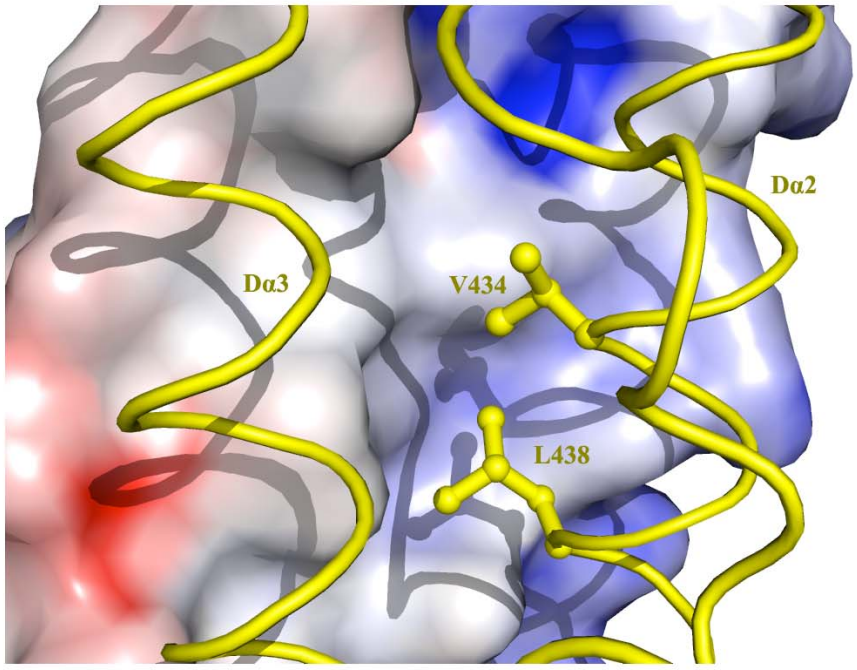
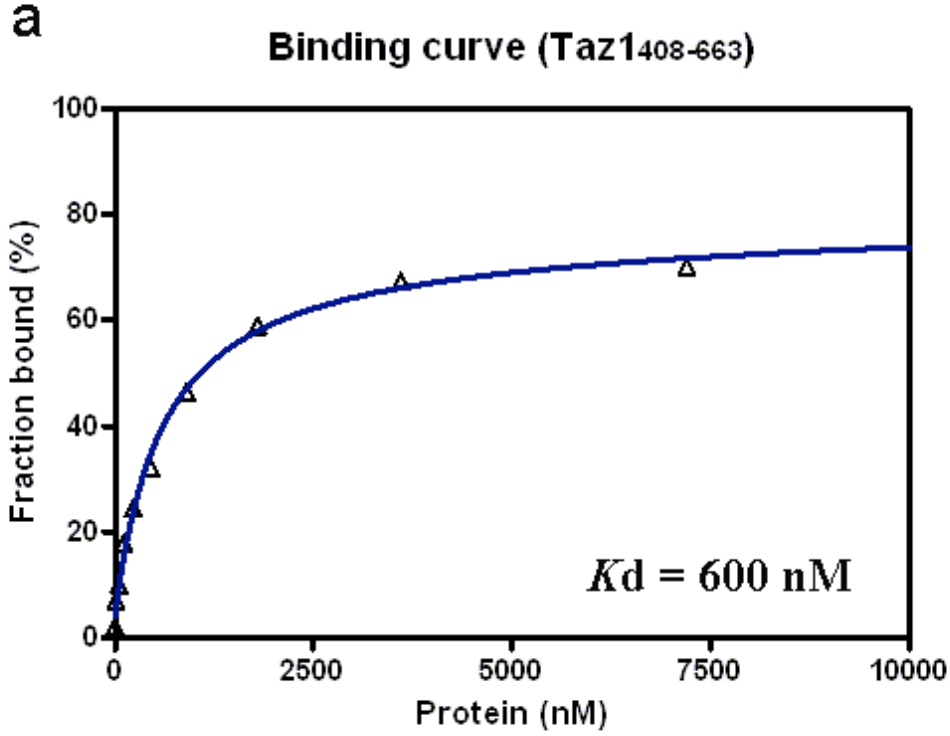

Dα2
Dα3
V434
L438

Figure 3.9 Disruption of Taz1 dimerization abolishes Taz1's dsDNA binding activity (a), (b) Equilibrium binding curves for the proteins of $Taz1_{408-663}$ wide type or L445R mutant. The dsDNA primer contains three pombe telomeric repeats "GGTTACA". The blue and red lines represent theoretical binding curves fit to the data for wild type and L445R mutant protein, respectively. The calculated equilibrium dissociation constant (*K*d) values are indicated.

a


Binding curve (Taz1408-663)
Fraction bound (%)
100
80
60
40
20
0
0
2500
5000
7500
10000
Protein (nM)
Kd = 600 nM

b

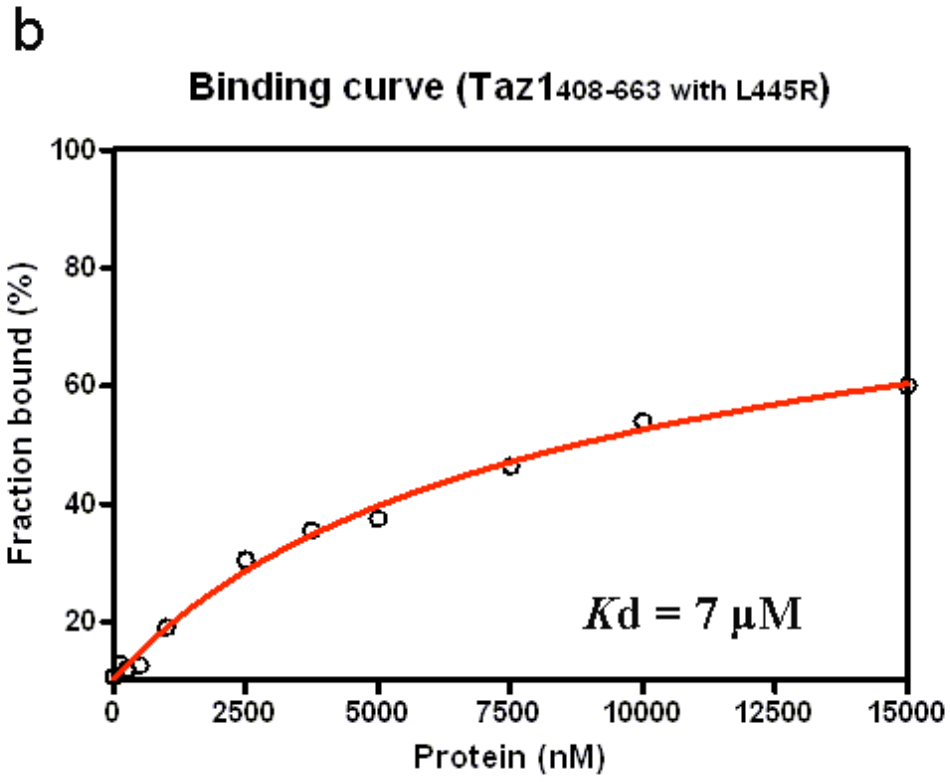

Binding curve (Taz1408-663 with L445R)
Fraction bound (%)
100
80
60
40
20
0
2500
5000
7500
10000
12500
15000
Protein (nM)
Kd = 7 µM

Figure 3.10 *In vivo* effects of the mutations on the Taz1 dimerization interface (Done by Dr. Julia Cooper's group.)
(a) Three Taz1 single mutations (L431R, V434W, and L445R) that showed no detectable dimerization ability in yeast two-hybrid assay exhibited significant loss of function in telomere length regulation as that in *Taz1Δ* cells (compare lane 2 with lane 3, 4 and 6). L438W mutant that retained partial dimerization ability in the yeast two-hybrid assay also retained partial function in suppressing telomere elongation.
(b) Two Taz1 mutants (L431R, L445R) and wild type are tagged. By ChIP, mutants L431R and L445R do not bind telomeric dsDNA (lower panel). On the dot blot (upper panel) of WCE (whole cell extract) that both mutants give a stronger signal with the TELO probe, indicating longer telomeres.
(c) Quantitative PCR is applied to monitor the telomere proximal position by using TELO primers, suggesting Taz1 mutants (L431R, L445R) do not bind telomeric DNA.

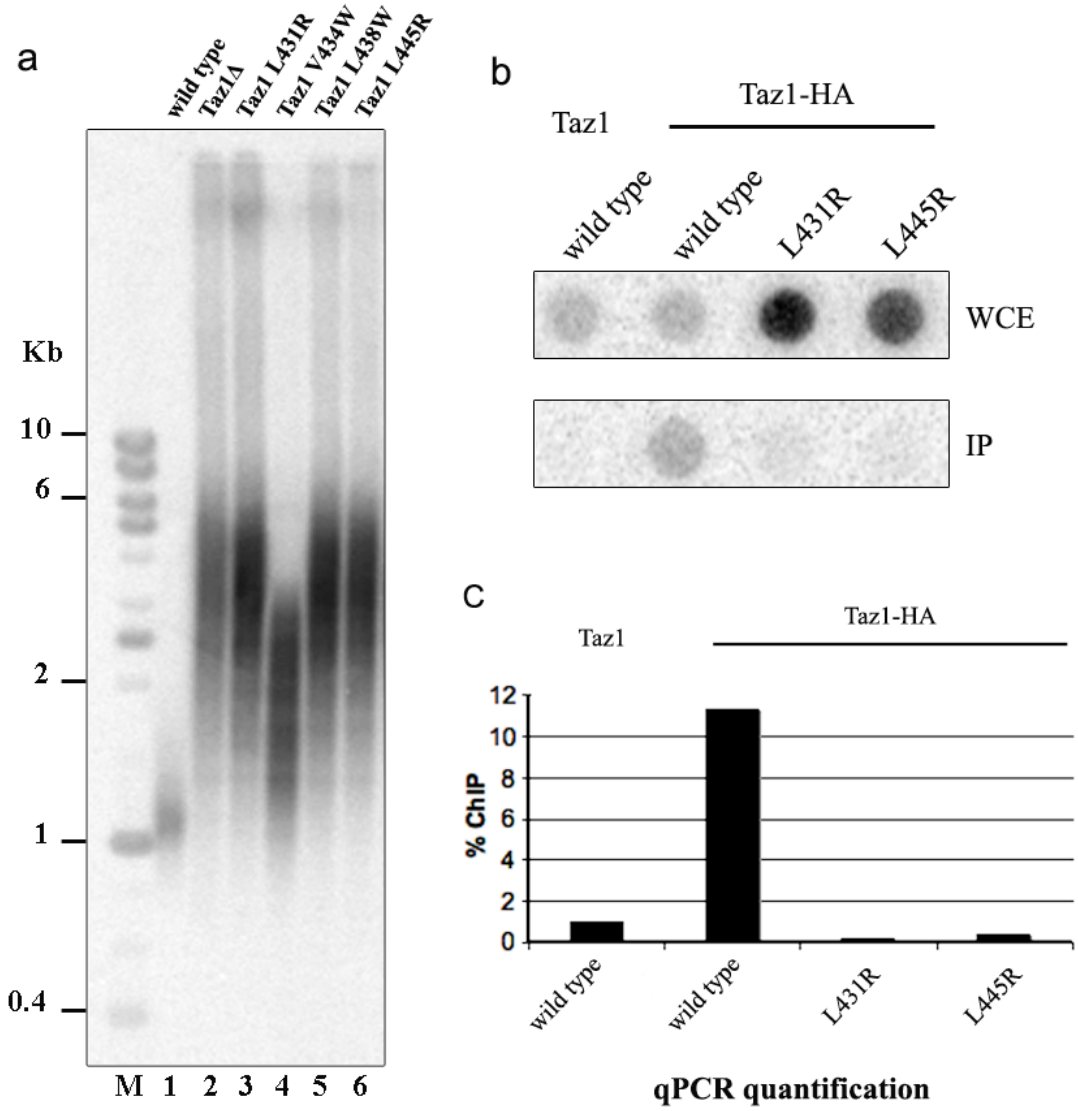
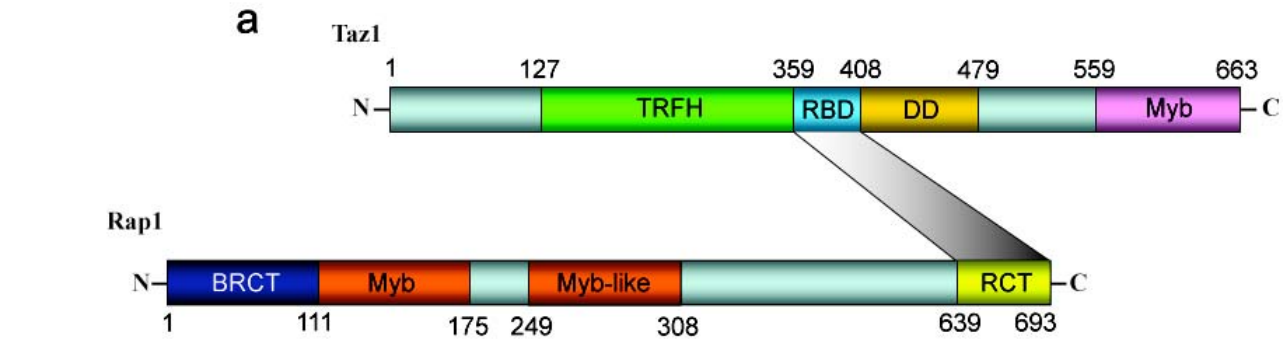
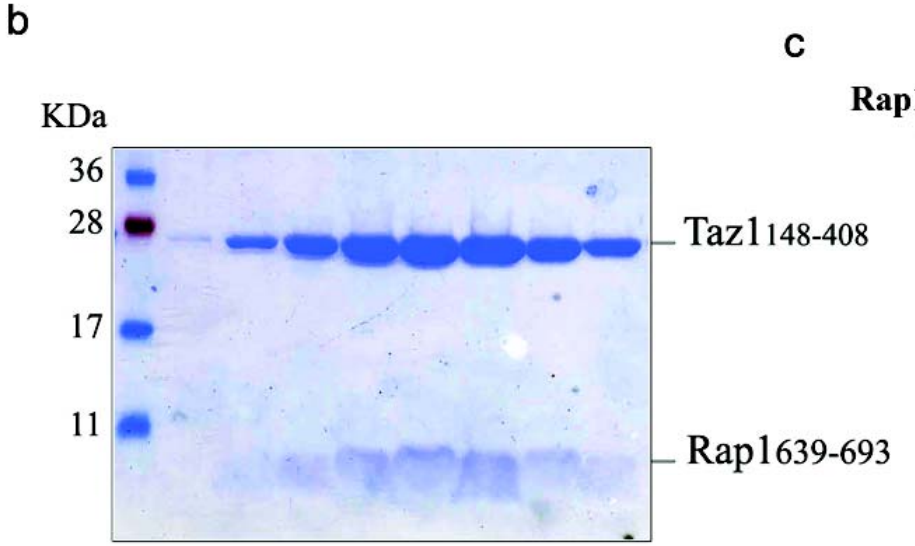

a
wild type
Taz1Δ
Taz1 L431R
Taz1 V434W
Taz1 L438W
Taz1 L445R
Kb
10
6
2
1
0.4
M 1 2 3 4 5 6
b
Taz1
Taz1-HA
wild type
wild type
L431R
L445R
WCE
IP
C
Taz1
Taz1-HA
% ChIP
12
10
8
6
4
2
0
wild type
wild type
L431R
L445R
qPCR quantification

Figure 3.11 Fusion protein of $Rap1_{RCT}$ - $Taz1_{RBD}$ complex
(a) Schematic illustration of domain organization of Taz1 and Rap1. $Taz1_{RBD}$ interacts with $Rap1_{RCT}$ as indicated.
(b) SDS-PAGE of $Taz1_{148-408}$-$Rap1_{639-693}$ complex.
(b) $Rap1_{RCT}$ and $Taz1_{RBD}$ was linked by a 14-residue peptide linker. The purified fusion protein of $Rap1_{RCT}$-$Taz1_{RBD}$ complex was shown in SDS-PAGE.


a
Taz1
1
127
359
408
479
559
663
N
TRFH
RBD
DD
Myb
C
Rap1
N
BRCT
Myb
Myb-like
RCT
C
1
111
175
249
308
639
693


b
KDa
36
28
17
11
Taz1148-408
Rap1639-693

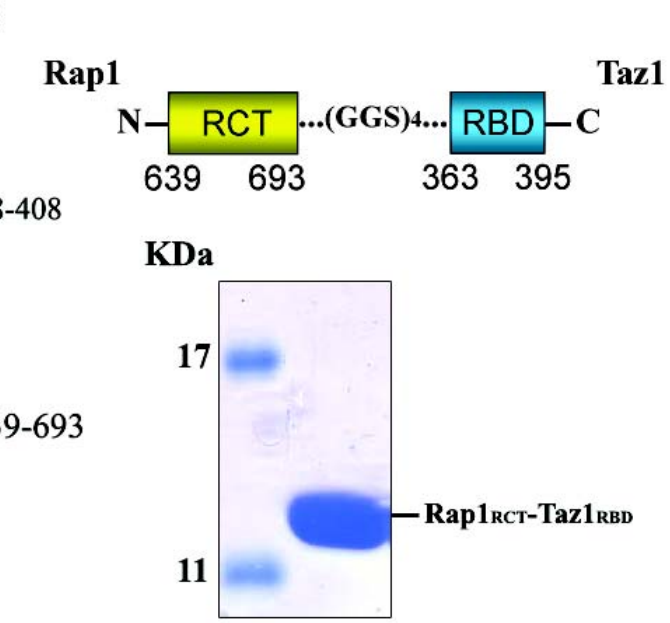

c
Rap1
Taz1
N
RCT
...(GGS)4...
RBD
C
639
693
363
395
KDa
17
11
Rap1RCT-Taz1RBD

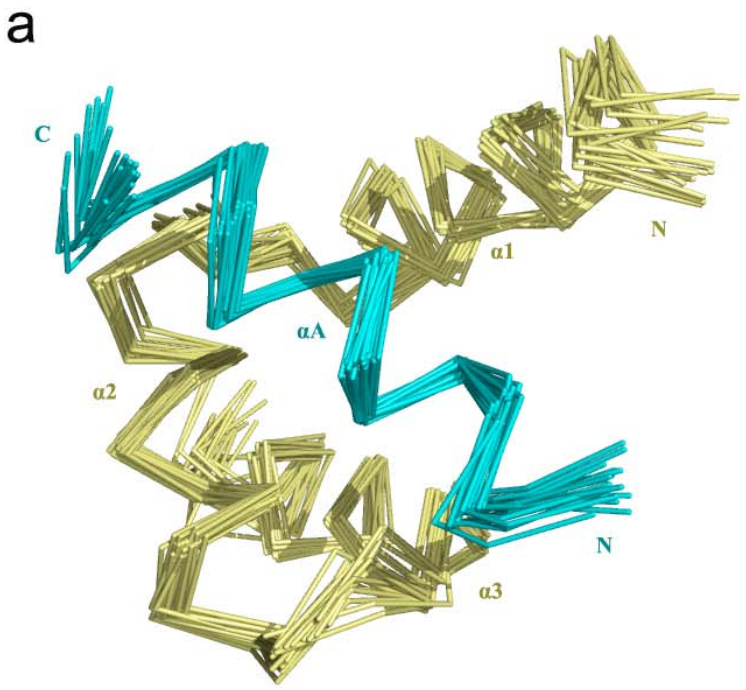
Figure 3.12 NMR structure of $Rap1_{RCT}$-$Taz1_{RBD}$ complex (Done by Dr. Hongyu Hu's group)

(a) Superposition of 20 energy-minimized structures of the $Rap1_{RCT}$-$Taz1_{RBD}$ complex. Rap1 (yellow) contains three α helices and Taz1 (cyan) contains one α helix.

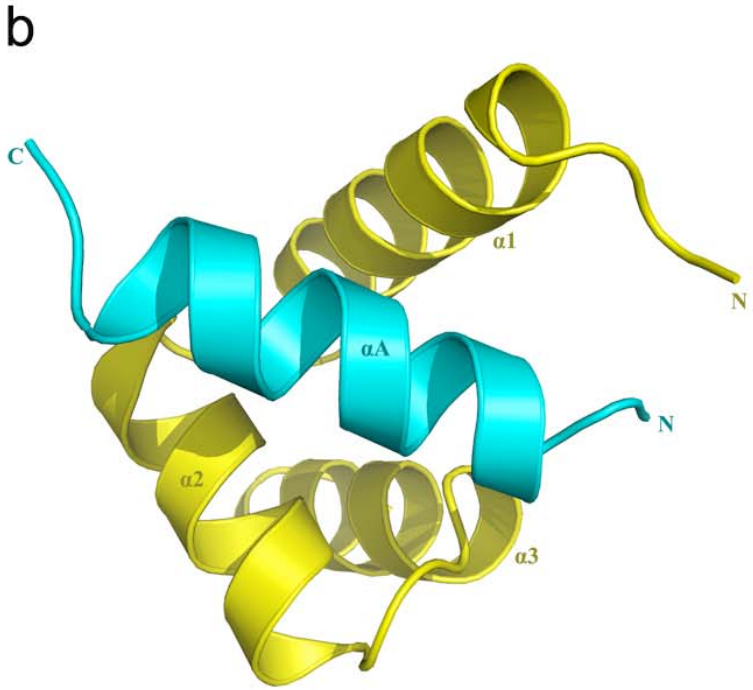
(b) Overall structure of $Rap1_{RCT}$-$Taz1_{RBD}$ complex.

(c), (d) $Rap1_{RCT}$-$Taz1_{RBD}$ interaction interface. $Rap1_{RCT}$ in (d) is in surface representation and colored according to its electrostatic potential (positive potential, blue; negative potential, red). Four hydrophobic residues on helix αA of $Taz1_{RBD}$ (Ile379, Leu380, Leu383 and Val387) wedge in between the two α-helices of $Rap1_{RCT}$: α1 and α2. Arg386 was deeply buried into a negative charged pocket of $Rap1_{RCT}$, which is made up by Asp652, Asp656 on α1 and Glu666 on α2. Arg384 forms two strong hydrogen bonds with the side chain of Glu674 on helix α2 of $Rap1_{RCT}$. Arg384 and Arg386 function as the two arms of a clamp to fix $Taz1_{RBD}$ on $Rap1_{RCT}$.

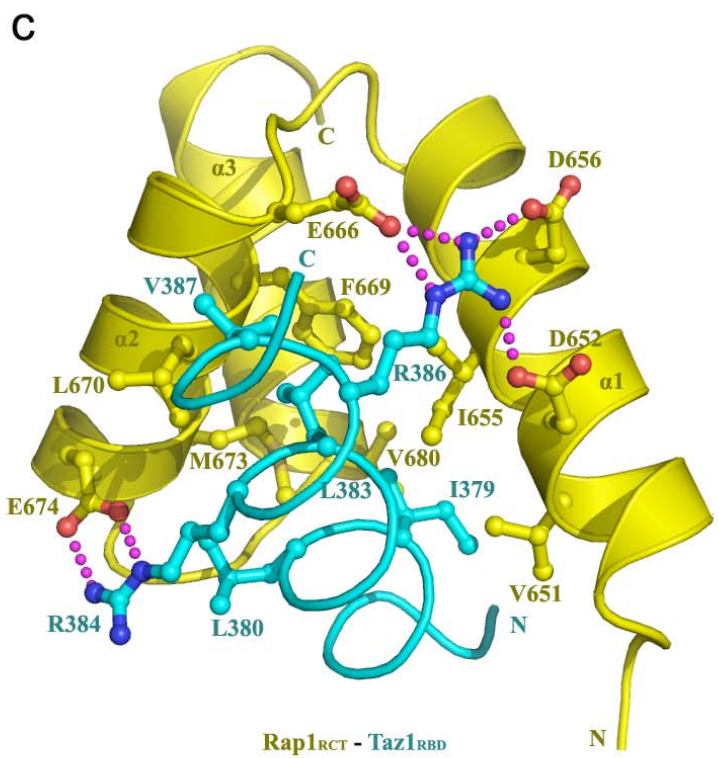
(e) $Rap1_{RCT}$ is in surface representation and colored according to its electrostatic potential (positive potential, blue; negative potential, red). The side chains of Ile379 and Leu383 of $Taz1_{RBD}$ are deeply buried in a hydrophobic groove made of a group of hydrophobic residues that locating in the center of $Rap1_{RCT}$ fold.

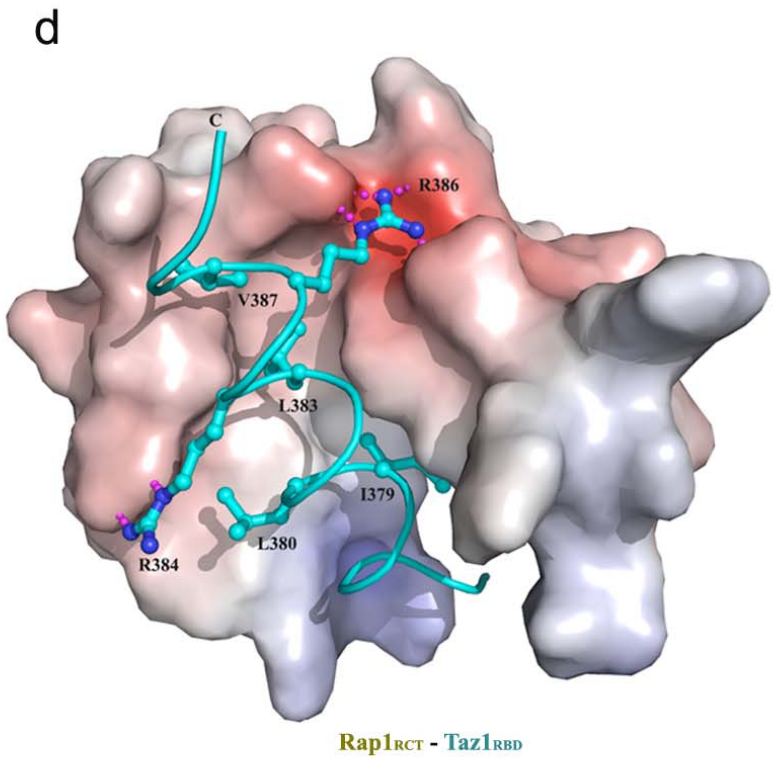


Rap1RCT - Taz1RBD



Rap1RCT - Taz1RBD


c
C
α3
D656
E666
C
V387
F669
α2
D652
R386
α1
L670
I655
M673
V680
L383
I379
E674
V651
R384
L380
N
N
Rap1RCT - Taz1RBD


d
C
R386
V387
L383
I379
L380
R384
Rap1RCT - Taz1RBD

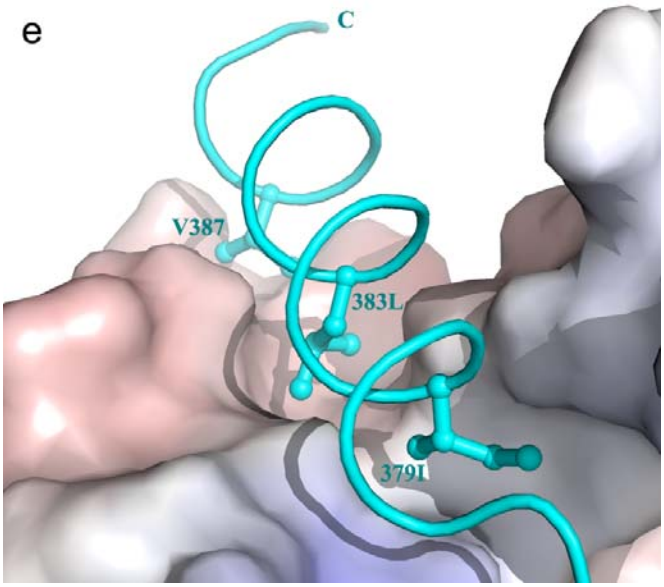

e
C
V387
383L
379I

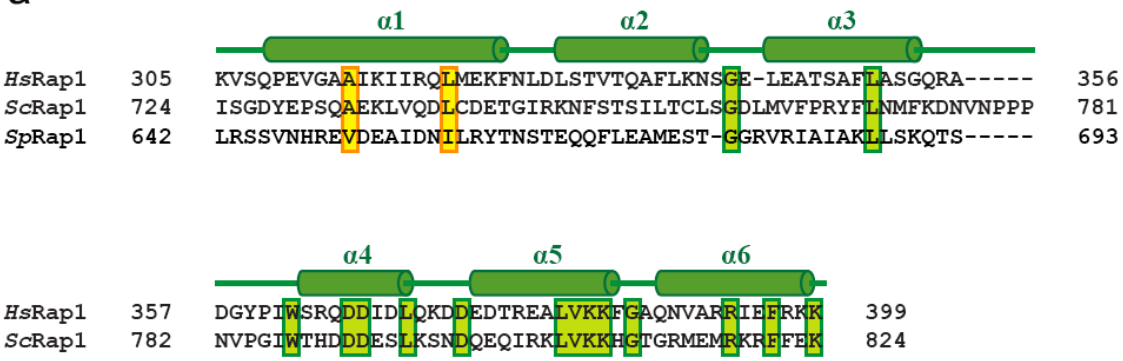
Table 3.3 Statistics for the structure of $Rap1_{RCT}$-$Taz1_{RBD}$ complex

No. of experimental restraints	
Total unambiguous distance restraints	1044
Intra-residual	521
Sequential (|i-j| = 1)	224
Medium ($2 \leq$|i-j| ≤ 4)	157
Long range (|i-j| ≥ 5)	142
Total ambiguous distance restraints	268
Hydrogen bond restraints	31
Dihedral angle restraints	
Φ	54
Ψ	54
Structure model statistics	
Coordinate precisions (Å) [a], residues 11-52, 85-97 (1-104)	
Backbone (N, C_α, CO)	0.51 ± 0.10 (4.85 ± 0.86)
Heavy atoms	1.46 ± 0.18 (5.05 ± 0.76)
RMSD from experimental restraints	
NOE distances (Å)	0.0834 ± 0.0173
Dihedral angles (deg.)	0.7398 ± 0.1493
RMSD from idealized geometry	
Bond lengths (Å)	0.0051 ± 0.0002
Bond angles (deg.)	0.6274 ± 0.0262
Impropers	1.6428 ± 0.1593
Energies (kcal/mol)	
E_{total}	-3711.0 ± 114.8
Restraint violations	
Distances (> 0.2 Å)	10
Dihedral angles (> 5 deg.)	0
Hydrogen bonds	0
Ramachandran analysis[b], residues 11-52, 85-97 (1-104)	
Residues in most favored regions	97.5 (75.8)
Residues in additionally allowed regions	2.5 (19.6)
Residues in generously allowed regions	0.0 (2.5)
Residues in disallowed regions	0.0 (2.1)

[a]The coordinate precision is defined as the average RMS deviation between the 20 structures and their mean coordinate positions.
[b]Exception of Gly residues.

Figure 3.13 Comparison of *Sp*Rap1$_{RCT}$, *Sc*Rap1$_{RCT}$ and human Rap1$_{RCT}$
(a) Sequence alignment of the RCT domains of human Rap1, *Sc*Rap1 and *Sp*Rap1. The alignment is based on the crystal structures of *Sc*Rap1$_{RCT}$ and human Rap1$_{RCT}$ and solution structure of *Sp*Rap1$_{RCT}$. *Sp*Rap1 lacks the last three helices (α4, α5 and α6). The identical residues are highlighted by green and the similar residues are highlighted by yellow.
(b) Superposition of the crystal structure of pombe Rap1$_{RCT}$-Taz1$_{RBD}$ complex and human TRF2$_{RBD}$-Rap1$_{RCT}$ complex (Chen *et al,* 2009, submitted). Helices are shown as colored cylinders; *Sp*Rap1$_{RCT}$ in yellow, Taz1$_{RBD}$ in cyan; human TRF2$_{RBD}$ in blue and Rap1$_{RCT}$ in red.

a

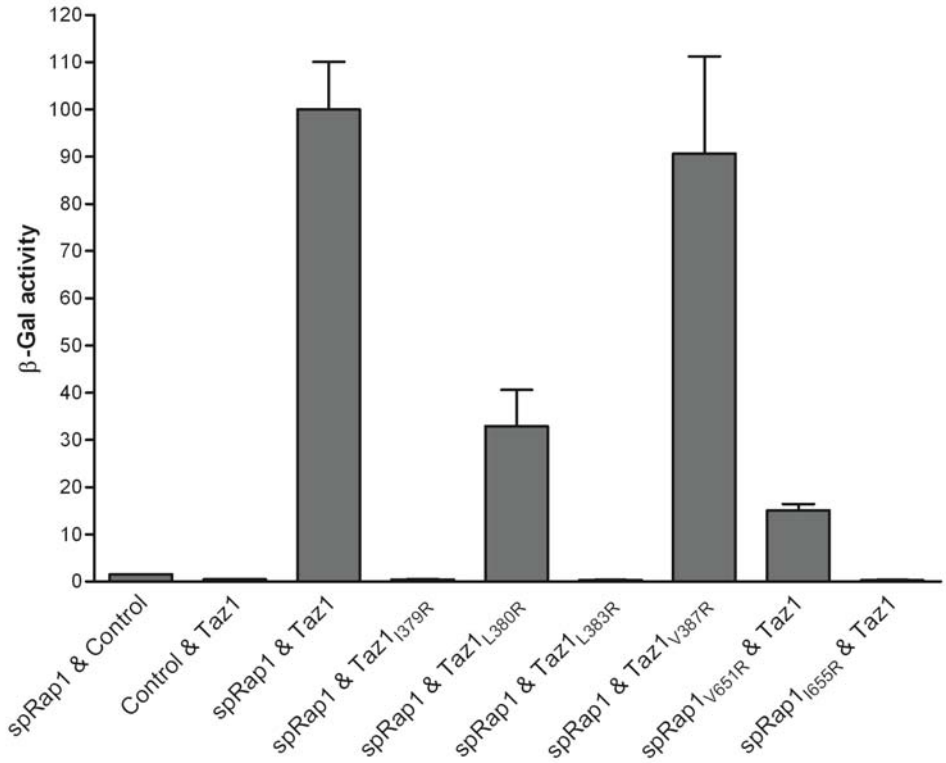

α1
α2
α3
HsRap1 305 KVSQPEVGAAIKIIRQLMEKFNLDLSTVTQAFLKNSGE-LEATSAFLASGQRA----- 356
ScRap1 724 ISGDYEPSQAEKLVQDLCDETGIRKNFSTSILTCLSGDLMVFPRYFLNMFKDNVNPPP 781
SpRap1 642 LRSSVNHREVDEAIDNILRYTNSTEQQFLEAMEST-GGRVRIAIAKLLSKQTS----- 693
α4
α5
α6
HsRap1 357 DGYPIWSRQDDIDLQKDDEDTREALVKKFGAQNVARRIEFRKK 399
ScRap1 782 NVPGIWTHDDDESLKSNDQEQIRKLVKKHGTGRMEMRKRFFEK 824

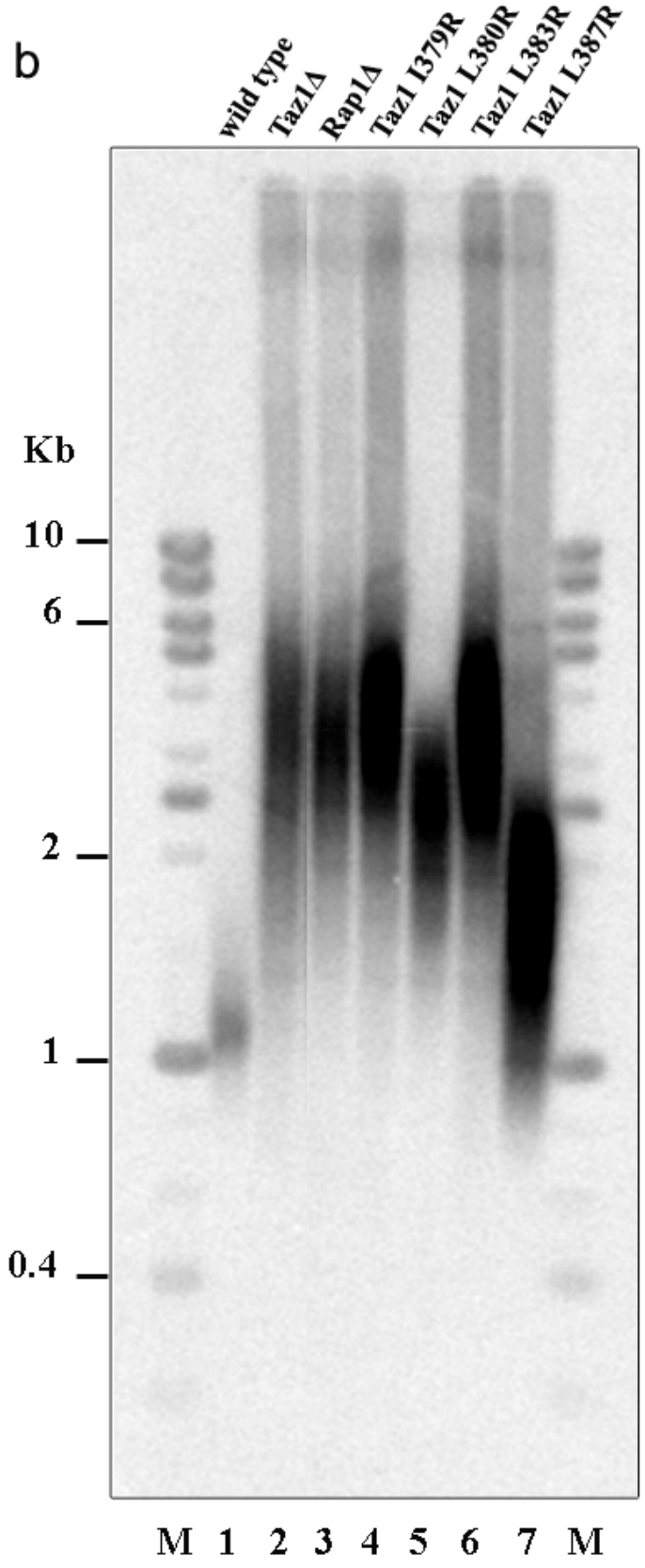
b


Rap1RCT - Taz1RBD / TRF2RBD - Rap1RCT
α1
C
N
αA
C
α3
αB'
α2
α6'
α4'
α5'

Figure 3.14 Effects of the mutations on the $Rap1_{RCT}$ - $Taz1_{RBD}$ interaction interface.
(a) In yeast two-hybrid, interaction of LexA-$Rap1_{RCT}$ with GAD-$Taz1_{RBD}$ was measured as β-galactosidase activity. Data are averaged of three independent β-galactosidase measurements.
(b) *In vivo* roles of the Taz1-Rap1 interaction in telomere length regulation. All four point mutants of Taz1 ($Taz1_{I379R}$, $Taz1_{L380R}$, $Taz1_{L383R}$, and $Taz1_{L387R}$) exhibited partial to complete loss of telomere length regulation, in a manner that is entirely consistent with the severity of the Rap1-binding defect. (Done by Dr. Julia Cooper's group.)

a


β-Gal activity
120
110
100
90
80
70
60
50
40
30
20
10
0
spRap1 & Control
Control & Taz1
spRap1 & Taz1
spRap1 & Taz1I379R
spRap1 & Taz1L380R
spRap1 & Taz1L383R
spRap1 & Taz1V387R
spRap1V651R & Taz1
spRap1I655R & Taz1


b
wild type
Taz1Δ
Rap1Δ
Taz1 I379R
Taz1 L380R
Taz1 L383R
Taz1 L387R
Kb
10
6
2
1
0.4
M 1 2 3 4 5 6 7 M

Reference:

1. Miyoshi, T., et al., *Fission yeast Pot1-Tpp1 protects telomeres and regulates telomere length.* Science, 2008. **320**(5881): p. 1341-4.
2. Cooper, J.P., et al., *Regulation of telomere length and function by a Myb-domain protein in fission yeast.* Nature, 1997. **385**(6618): p. 744-7.
3. Li, B., S. Oestreich, and T. de Lange, *Identification of human Rap1: implications for telomere evolution.* Cell, 2000. **101**(5): p. 471-83.
4. Fairall, L., et al., *Structure of the TRFH dimerization domain of the human telomeric proteins TRF1 and TRF2.* Mol Cell, 2001. **8**(2): p. 351-61.
5. Tomita, K., et al., *Competition between the Rad50 complex and the Ku heterodimer reveals a role for Exo1 in processing double-strand breaks but not telomeres.* Mol Cell Biol, 2003. **23**(15): p. 5186-97.
6. Miller, K.M., M.G. Ferreira, and J.P. Cooper, *Taz1, Rap1 and Rif1 act both interdependently and independently to maintain telomeres.* EMBO J, 2005. **24**(17): p. 3128-35.
7. Kanoh, J. and F. Ishikawa, *spRap1 and spRif1, recruited to telomeres by Taz1, are essential for telomere function in fission yeast.* Curr Biol, 2001. **11**(20): p. 1624-30.
8. Chikashige, Y. and Y. Hiraoka, *Telomere binding of the Rap1 protein is required for meiosis in fission yeast.* Curr Biol, 2001. **11**(20): p. 1618-23.
9. Marcand, S., E. Gilson, and D. Shore, *A protein-counting mechanism for telomere length regulation in yeast.* Science, 1997. **275**(5302): p. 986-90.
10. Teixeira, M.T., et al., *Telomere length homeostasis is achieved via a switch between telomerase- extendible and -nonextendible states.* Cell, 2004. **117**(3): p. 323-35.
11. van Steensel, B. and T. de Lange, *Control of telomere length by the human telomeric protein TRF1.* Nature, 1997. **385**(6618): p. 740-3.
12. Spink, K.G., R.J. Evans, and A. Chambers, *Sequence-specific binding of Taz1p dimers to fission yeast telomeric DNA.* Nucleic Acids Res, 2000. **28**(2): p. 527-33.
13. Li, B., A. Espinal, and G.A. Cross, *Trypanosome telomeres are protected by a homologue of mammalian TRF2.* Mol Cell Biol, 2005. **25**(12): p. 5011-21.
14. Gao, H., et al., *RPA-like proteins mediate yeast telomere function.* Nat Struct Mol Biol, 2007. **14**(3): p. 208-14.
15. Lupas, A., *Coiled coils: new structures and new functions.* Trends Biochem Sci, 1996. **21**(10): p. 375-82.
16. Burkhard, P., J. Stetefeld, and S.V. Strelkov, *Coiled coils: a highly versatile protein folding motif.* Trends Cell Biol, 2001. **11**(2): p. 82-8.
17. Mason, J.M. and K.M. Arndt, *Coiled coil domains: stability, specificity, and biological implications.* Chembiochem, 2004. **5**(2): p. 170-6.
18. Bianchi, A., et al., *TRF1 is a dimer and bends telomeric DNA.* EMBO J, 1997. **16**(7): p. 1785-94.
19. Nishikawa, T., et al., *Solution structure of the DNA-binding domain of human telomeric protein, hTRF1.* Structure, 1998. **6**(8): p. 1057-65.

20. Holm, L. and C. Sander, *Database algorithm for generating protein backbone and side-chain co-ordinates from a C alpha trace application to model building and detection of co-ordinate errors.* J Mol Biol, 1991. **218**(1): p. 183-94.
21. Miller, K.M., O. Rog, and J.P. Cooper, *Semi-conservative DNA replication through telomeres requires Taz1.* Nature, 2006. **440**(7085): p. 824-8.
22. Sfeir, A., et al., *Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication.* Cell, 2009. **138**(1): p. 90-103.
23. Broccoli, D., et al., *Comparison of the human and mouse genes encoding the telomeric protein, TRF1: chromosomal localization, expression and conserved protein domains.* Hum Mol Genet, 1997. **6**(1): p. 69-76.
24. Wang, F., et al., *The POT1-TPP1 telomere complex is a telomerase processivity factor.* Nature, 2007. **445**(7127): p. 506-10.
25. Brunger, A.T., et al., *Crystallography & NMR system: A new software suite for macromolecular structure determination.* Acta Crystallogr D Biol Crystallogr, 1998. **54**(Pt 5): p. 905-21.
26. Jones, T.A., et al., *Improved methods for building protein models in electron density maps and the location of errors in these models.* Acta Crystallogr A, 1991. **47 (Pt 2)**: p. 110-9.
27. Delaglio, F., et al., *NMRPipe: a multidimensional spectral processing system based on UNIX pipes.* J Biomol NMR, 1995. **6**(3): p. 277-93.
28. Rieping, W., et al., *ARIA2: automated NOE assignment and data integration in NMR structure calculation.* Bioinformatics, 2007. **23**(3): p. 381-2.
29. Duggan, B.M., et al., *SANE (Structure Assisted NOE Evaluation): an automated model-based approach for NOE assignment.* J Biomol NMR, 2001. **19**(4): p. 321-9.
30. Guntert, P., C. Mumenthaler, and K. Wuthrich, *Torsion angle dynamics for NMR structure calculation with the new program DYANA.* J Mol Biol, 1997. **273**(1): p. 283-98.
31. Cornilescu, G., F. Delaglio, and A. Bax, *Protein backbone angle restraints from searching a database for chemical shift and sequence homology.* J Biomol NMR, 1999. **13**(3): p. 289-302.
32. Wishart, D.S. and B.D. Sykes, *The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data.* J Biomol NMR, 1994. **4**(2): p. 171-80.
33. Laskowski, R.A., et al., *AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR.* J Biomol NMR, 1996. **8**(4): p. 477-86.
34. Cavalier-Smith, T., *Palindromic base sequences and replication of eukaryote chromosome ends.* Nature, 1974. **250**(5466): p. 467-70.
35. Chen, Y., et al., *A shared docking motif in TRF1 and TRF2 used for differential recruitment of telomeric proteins.* Science, 2008. **319**(5866): p. 1092-6.

CHAPTER 4

STRUCTURAL ANALYSES OF BLM-TOPOIIIα-RMI1-RMI2 COMPLEX, THE DOUBLE HOLLIDAY JUNCTION DISSOLVASOME.

The BLM-TOPOIIIα-RMI1-RMI2 (BTR) complex plays an important role in dissoluting homologous recombination (HR) intermediate double Holliday junctions (dHJs). Recent data suggest that the dHJ dissolvasome BTR complex is also involved in telomere maintenance in telomerase-negative ALT (alternative lengthening of telomeres) cells. However, the detailed dissolution mechanism of the BTR complex is unclear. In this chapter, I focus on the structural and functional characterization of the RMI1-RMI2 complex followed by discussing on the studies towards the structures of the BTR complex.

4.1 Abstract

Telomeres in most cancer cells and tumor tissues are maintained by telomerase for long-term proliferation. While in some types of cancer cells, telomeres could be maintained by a mechanism named ALT (alternative lengthening of telomeres) in the absence of telomerase. BLM, the RecQ helicase mutated in the Bloom Syndrome, is found in a special nuclear structure called ALT-associated PML (promyelocytic leukemia) bodies (APBs). BLM associates with TOPOIIIα and RMI1-RMI2 to form the BTR (BLM-

TOPOIIIα-RMI1-RMI2) complex. This complex can specifically resolve the double Holliday junction (dHJ), a homologous recombination (HR) intermediate, without crossover recombinants. The BTR complex may also function in telomere maintenance in ALT cells because 1) telomeres in ALT cells could be processed through the HR pathway, and 2) BLM and double-stranded telomeric DNA binding proteins TRF1 and TRF2 are found in APBs. Direct interactions between BLM and TRF1/TRF2 have been reported, and 3) the BTR complex associates human telomeric chromatin from ALT cells. Collectively, the dHJ dissovalsome BTR complex may play an important role in telomere metabolism in ALT cells. In order to elucidate the structural basis of dHJ dissolution mechanism of the BTR complex, I first solved the crystal structure of the RMI1-RMI2 complex. The RMI1-RMI2 complex contains three OB folds (two in RMI1 and one in RMI2), which interact in a similar way as RPA1 and RPA2 in the RPA complex. Structural study of BLM was also performed. I have already successfully obtained crystals of the catalytic core of BLM. Functional interpretations based on the structural information are also tested in an *in vitro* dHJ dissolution assay.

4.2 Introduction

4.2.1 Alternative Lengthening of Telomere in Mammalian Cells

Telomerase plays an important role in cells of the germ line and in normal somatic biology, especially in those tissue compartments that depend on extensive cellular proliferation. Nevertheless, in normal somatic cells telomerase is not expressed at sufficient levels to prevent telomere shortening due to the end replication problem. In contrast, telomere length maintenance in most cancer cells requires dysregulated

telomerase activity [1]. A substantial minority of immortalized mammalian cell lines and tumors are telomerase-negative, however, and in these cells telomere length maintenance can be achieved instead by a telomerase-independent mechanism, called Alternative Lengthening of Telomeres (ALT) [2]. The existence of ALT was deduced from observations of telomere length maintenance over many hundreds of population doublings in the absence of detectable telomerase activity.

ALT-positive human cells can be recognized on the basis of a number of hallmarks. One unique feature of ALT cell lines and tumors is the presence of special nuclear structure called ALT-associated promyelocytic leukemia (PML) bodies (APBs). PML bodies are found in several different cellular processes including cell cycle regulation, senescence, apoptosis, tumor suppression, and immune and inflammatory responses [3-8] (reviewed in [4]). APBs appear to be special since in addition to recombination proteins RAD52, RAD51, RPA, RecQ helicase WRN and BLM, they also contain the telomeric DNA and duplex telomeric DNA binding proteins TRF1 and TRF2 [9]. ALT cells can also be distinguished from other cells by the extremely heterogeneous telomere length, the increased number of extra-chromosomal telomeric DNA repeat (ECTR) circles (or T-circles) and dramatic telomere lengthening and shortening events [10-14]. Under normal conditions, cultured human somatic cells lose about 40 – 200 bp during each cell cycle [15-18]. In human germ line cells, the telomere length is maintained at the level of ~ 15 kb by telomerase [19, 20], and in human cancer cells the telomeres are normally homogeneous and the average length is below ~ 10 kb [20-22]. In contrast, telomeres in ALT cells are very heterogeneous, ranging from < 3 kb to > 50 kb with the average length of ~ 20 kb [2, 10, 11, 23, 24].

The rapid dynamics of telomere length and the increased number of T-circles in ALT cells suggest that telomere maintenance mechanism in these cells involves homologues recombination (HR). HR is a conserved mechanism in which the exchange of genetic information (nucleotide sequences) between two similar (homologous) or identical strands of DNA sequences. During the process of HR, several HR intermediates appear, including D-loops and double Holliday Junctions (dHJs). Coincidently, the T-loop, the reshaped telomere structure by the shelterin complex, resembles an intermediate of HR (see Chapter 1, Figure 1.10), and provides a substrate for telomere deletion through the HR pathway. Previous studies showed that the N-terminal basic domain (B-domain) of TRF2 prevents the T-loop from undergoing HR mediated deletion [25]. Expression of TRF2ΔB in cells induced catastrophic deletions of telomeric DNA, and the number of T-circles increased to a level similar to that in ALT cells [25]. These T-circles are very likely the resolution products of T-loops as electron microscopy data showed that these T-circles and T-loops were of similar size[25, 26] (Fig. 4.1).

4.2.2 The BLM-TOPOIIIα-RMI1-RMI2 (BTR) complex, a dHJ dissolvasome that associated with telomeres

BLM belongs to the RecQ helicase family, named after the *E.coli recQ* gene product. There are five RecQ helicase members in humans: RECQ1, RECQ4, RECQ5, WRN and BLM [27-32]. Defects in three RecQ helicases cause genome instability and premature aging [32]. Specifically, defects in *BLM* gene result in cancer predisposition disorders BS (Bloom's syndrome) [33]. BS is an autosomal recessive genetic disorder in human. BS patients have a very high risk of most types of cancers at the early age of onset (usually at

about 24 years of age). It has been identified that the mutated gene (*BLM*) in BS patients locates on chromosome 15q26.1 [33, 34]. The *BLM* gene encodes a protein of 1417 amino acids, which is expressed in all active proliferation tissues, such as testis, ovary, thymus and spleen [32, 35]. There is a high BLM protein expression level in S phase and it reaches a peak in G2 phase [36, 37].

As a member of the RecQ helicase family, BLM is able to unwind various HR intermediates *in vitro*, including D-loops and Holliday Junctions [38-40]. It has been observed experimentally that in human cells BLM functionally associates with a type IA topoisomerase, topoisomerase IIIα (TOPOIIIα), which unknots the two DNA strands by introducing transient nicks [41, 42]. Through the combined and cooperative nicking and unwinding activities, one particular HR intermediate, known as a double Holliday Junction (dHJ) that arises from reciprocal exchanges of single stranded sequence during the process of HR, is processed by the BLM complex without leaving any crossover recombinants, in a process termed 'double Holliday junction disolution' (Fig. 4.2) [43-46](reviewed in [47]). Thus, the BLM complex prevents the formation of hybrid recombinant DNA molecules called crossovers that could lead to a phenomenon known as loss of heterozygosity, which significantly contributes to cancer.

Under physiological conditions, the dissolution activity of BLM-TOPOIIIα requires a third member of the BLM complex, RMI1 (RecQ mediated genome instability 1 or BLAP75) [46, 48]. Since RMI1 has ssDNA binding ability, it has been proposed that RMI1 acts as an accessory factor to load TOPOIIIα onto the substrate and stimulate TOPOIIIα activity in dHJ dissolution [43, 49]. Recently, a novel component of the dissolvasome complex, RMI2 (or BLAP18), was identified [50, 51]. RMI2 helps

maintain the stability of the BLM-TOPOIIIα-RMI1 complex and plays a key role in mitotic phosphorylation of BLM and suppressing sister-chromatid exchange (SCE) [50]. Through interacting with the carboxyl-terminus of RMI1, RMI2 associates with BLM-TOPOIIIα-RMI1 to form a stable BLM-TOPOIIIα-RMI1-RMI2 (BTR) complex (also called the "dissolvasome") [50, 51]. Notably, TOPOIIIα, RMI1 and RMI2 could also form a stable protein complex *in vivo* without BLM. Deletion of any one of the three components by RNA interference will result in the disruption of the complex, while the BLM expression level is unaffected [50, 51].

Several lines of evidence suggested that the BTR complex is required for telomere maintenance in ALT cells. First, both BLM and telomeric DNA were found in APBs [9]. Second, the telomeric DNA structure T-loop resembles the HR intermediates D-loop and Holliday junction, which are the substrates of the BTR complex (Fig. 4.1). Third, it has been shown that BLM directly interacts with telomere binding protein TRF2, indicating a connection between BLM and telomere maintenance [52-54]. Last but not least, a newly developed technique called proteomics of isolated chromatin segments (PICh) revealed that all four components of the BTR complex were found to associate with human telomeric chromatin in ALT cells, but not in telomerase-positive cancer cells [55]. Therefore, we proposed that the dHJ dissolvasome BTR complex plays a role in telomere maintenance in the absence of telomerase. The detailed molecular mechanism of dHJ dissolution catalyzed by BTR remains unclear. In order to address this question and elucidate the potential functions of the BTR complex at telomeres, I set up a project to study the structure of the BTR complex.

4.3 Structural determination of the RMI1-RMI2 complex

Sequence alignment analysis predicted that both RMI1and RMI2 are OB-fold-containing proteins [51]. RMI1 has two putative OB-folds at both termini and RMI2 has one. The C-terminal OB-fold of RMI1 (RMI1C) shares limited similarity to the C-terminal OB-fold of RPA1 (RPA1C), the largest subunit of the RPA complex. The OB-fold in RMI2 is most similar to the OB-fold of another RPA protein, RPA2. However, the sequence of the N-terminal OB-fold of RMI1 (RMI1N) shows no apparent similarity to any OB-fold in the RPA complex [50]. RPA is a three-subunit complex that binds single-stranded DNA and plays an essential role in DNA replication, DNA recombination, and DNA repair [56-58]. RPA contains six OB-folds in total: four in RPA1 (RPA1-F, RPA1-A, RPA1-B, RPA1-C), one in each of RPA2 and RPA3 [56-58].

There is a 280-residue unstructured fragment between the two OB-folds of RMI1, which has no secondary structural feature based on secondary structural prediction analysis. Consistently, initial attempt to express full-length RMI1 yielded no soluble protein. Therefore, two fragments of RMI1 that correspond to the two OB-folds, RMI1N (residues 2 – 213) and RMI1C (residues 475 – 625), were cloned separately into the His6-SUMO vector and both proteins were expressed and purified from *E. coli.*

The RMI1N construct produced soluble and stable proteins, which allowed me to obtain enough material of high purity for crystallization trails (Fig. 4.3a-c). Thin plate crystal clusters started to show up two days after the initial screening. After several rounds of optimization, I obtained good quality crystals that are suitable for structural study (Fig. 4.3d). Diffraction data were collected at the Advanced Photon Source (APS). The space group is $P3_121$ with one molecule per asymmetric unit. The structure of

RMI1N was solved by single-wavelength anomalous dispersion (SAD) using Seleno-Met substituted crystals and refined to 2.0 Å resolution (Table 4.1).

In contrast to the N-terminal OB-fold of RMI1, the expression of RMI1C yielded insoluble protein. Since RMI2 interacts with the RMI1C, we reasoned that the correct folding and/or the stability of RMI1C might need the existence of RMI2. Then we cloned full-length RMI2 into the pGEX-6p-1 vector, and His6-SUMO-tagged RMI1C and GST-tagged RMI2 were co-expressed in Rosetta DE3 strain under a Kana/Amp double selection. The RMI1C-RMI2 complex is well expressed and kept in a soluble and mono-dispersed form. The heterodimeric interaction was strong enough that the RMI1C-RMI2 complex was able to tolerate several rounds of size exclusion and ion-exchange chromatographic steps, and still maintain a 1:1 stoichiometry. The RMI1C-RMI2 complex with > 99% purity was subjected to crystallization trial and chunky crystal clusters were obtained one week after initial screening (Fig. 4.4). High-resolution data sets were collected at APS and the phase information was obtained by diffracting Seleno-Met substituted derivative crystals. The structure of the RMI1C-RMI2 complex was solved by SAD method and refined to 1.9 Å resolution. The space group is $P2_12_12_1$ with one complex per asymmetric unit (Table 4.2).

4.4 Crystal structure of RMI1N

The crystal structure reveals that RMI1N indeed contains a single OB-fold core, as expected from secondary structural prediction results (Fig. 4.5a). N-terminal to the OB-fold core (residue 2 – 56) is a three-helix-bundle motif, which packs on the OB-fold though hydrophobic van der Waals contacts and buries a total surface area of 2453.8 Å

(Fig. 4.5c). This observation and our previous unsuccessful attempt to purify the RMI1 N-terminal OB-fold without the three-helix bundle suggest that this motif is crucial for the OB-fold stability (Section 4.3). Although bioinformatics analysis failed to detect any substantial sequence similarity between RMI1N and RPA proteins, comparison of the structures of RMI1N and the N-terminal OB-fold of RPA1 (RPA1Nl residue 1 – 120) clearly reveals a high degree of structural similarity (Fig. 4.5d). In fact, RPA1N is one of the top solutions revealed by Dali that are structurally most similar to RMI1N with an rmsd of 2.5 Å for 118 equivalent Cα atoms. This close structural similarity is rather unexpected given that the sequences of the OB folds of Ten1 and Rpa3 share only 15% identity.

Despite these similarities, the structure of RMI1N contains several unique features. Neither end of the RMI1N OB-fold β-barrel is covered by α-helix, which is often observed in OB-fold-containing structures, leaving a β-barrel with an open hole in the center (Fig. 4.5b). Notably, there is a large 35-residue insertion (residue 97 – 131) between strands 1 and 2 (Fig. 4.5a, b, d). This large insertion, which contains one α-helix and a long loop (partially disordered), extends out from the protein perpendicular to the axis of the OB-fold β-barrel. The substitution of the insertion by a four-residue-linker "SGGS" between 1 and 2 did not affect the protein stability and generated well-folded products (the mutant protein was named as RMI1N$^{\Delta loop}$). RMI1N$^{\Delta loop}$ will be used in various biochemical assays to investigate the function of this large insertion in RMI1. By collaborating with Dr. Patrick Sung's lab from Yale University, we test the RMI1N and RMI1N$^{\Delta loop}$ in pull down assays and *in vitro* dHJ dissolution assay. RMI1N interacts with both BLM and TOPOIIIα in pull down assay (Fig. 4.6a) and is active in dHJ dissolution

assay. With BLM- TOPOIIIα, RMI1N can dissolute 50% of the dHJ substrate, even higher than the RMI1 full-length protein (40% dissolution) at similar concentration (Fig. 4.6b). However, although RMI1N$^{\Delta loop}$ retains the binding ability to BLM and TOPOIIIα (Fig. 4.6a), it fails to enhance the dissolution activity of BLM-TOPOIIIα (Fig. 4.6b). Therefore, the N-terminal of RMI1 is responsible for stimulating BLM-TOPOIIIα and the loop region outside the OB-fold may play an essential role in dHJ dissolution. Further investigations need to be applied in order to elucidate more detailed information.

4.5 Crystal structure of the RMI1C-RMI2 complex

The RMI1C-RMI2 complex structure reveals a 1:1 stoichiometry between RMI1C and RMI2, consistent with the observed molecular weight of the complex as determined by gel filtration chromatography (~38 kDa, Fig. 4.4b, c). The crystal structure shows that each protein indeed comprises a single OB-fold (Fig. 4.7a). In addition to the central β-barrel, there are several structural features common to the OB folds of RMI1C and RMI2. First, both proteins contain a C-terminal α helix (α4 in RMI1C and α3' in RMI2), which contributes most of the contact interface between RMI1C and RMI2 (Fig. 4.7a, b). Second, short α helices (α3 in RMI1C, and α2' in RMI2) that cover the bottom of the β-barrels of the OB folds are found between strands β3 and β4 (Fig. 4.7a). Third, the other end of the β-barrel of both proteins is closed by α helix (α2 in RMI1C and α1' in RMI2) N-terminal to the OB folds (Fig. 4.7a). Notably, compared to RMI2, RMI1C contain a unique segment in the connecting region between β3 and β4. This segment, which consists of resides 558-587 of RMI1, folds into two helices β4 and β5, extending outside of the OB-fold core (Fig. 4.7a).

The interface between RMI1C and RMI2 in the crystal structure is mainly hydrophobic. The interactions are mediated primarily by the C-terminal amphipathic helices of both proteins; hydrophobic residues from RMI1C (Met613, Leu616 and Leu619) and RMI2 (Ile130, Met134 and Leu137) inter-digitate with one another to form the core of the hydrophobic interface (Fig. 4.7b). The α3' helix of RMI2 makes a 30° kink that bends the helix towards helix α4 of RMI1C, extending the interaction surface between the two OB folds. In addition to hydrophobic contacts, hydrogen-bonding interactions between the two C-terminal helices appear also to strengthen the interface and contribute to the specificity of the RMI1C-RMI2 complex. Specifically, at the C-terminal end of the RMI1C α6 helix, the side chain of R622 makes two salt-bridge interactions with the carboxylate group of D141 in the middle of RMI2 α3' helix (Fig. 4.7c). A panel of missense mutations designed to disrupt the interface have been generated and their effects on the RMI1-RMI2 interaction will be examined using yeast two-hybrid assay.

The crystal structure of the RMI1C-RMI2 complex closely resembles that of the RPA1C-RPA2N complex (Fig. 4.8a) [51]. Consistent with previous sequence alignment predictions, the structure of RMI1C and RNI2 OB-folds are most similar to those of the OB-folds of RPA1C and RPA2N, with rmsd values of 3.2 Å and 2.0 Å, respectively (Fig. 4.8b) [51]. Notably, the structurally conserved region includes not only the central β-barrel of the OB fold, but also peripheral α-helices including the three helices between strands β3 and β4 (Fig. 4.8a-c). In addition to similarities between the individual components, the RMI1C-RMI2 and the RPA1C-RPA2N complexes exhibit another common feature; in both cases, the two subunits heterodimerize mainly through

hydrophobic contacts mediated by the two C-terminal helices (Fig. 4.8a). Taken together, we concluded that RMI1C-RMI2 is structurally similar to the RPA1C-RPA2N complex.

Despite the high degree of overall structural conservation, there are substantial differences between the RMI1C-RMI2 and the RPA1C-RPA2N complexes. Most notably, the relative orientations between the two components are different in the two complexes. When both complex structures are overlaid based on the OB folds of RMI1C and RPA1C, RMI2 exhibits a ~ 16° rotation relative to the position of RPA2N (Fig. 4.8a). Second, RMI1C lacks a zinc-binding motif between strands β1 and β2 (Fig. 4.8a). Instead β1 and β2 are connected by a short seven-residue loop (residues 522-528). The zinc-binding motif in RPA1 is essential for the *in vivo* function of the RPA complex [59, 60]. Nevertheless, we conclude that the RMI1C-RMI2 complex is a structural homologue of RPA1C-RPA2.

The RPA complex contains three components: RPA1C, RPA2 and RPA3 [56-58]. Given the structural similarity between RMI1-RMI2 and RPA1-RPA2, I speculated that it is likely that there exists an unidentified protein that binds to RMI1-RMI2 to form an RPA-like trimer. Indeed, a potential candidate RMI1-RMI2 binding protein that reassembles RPA3 has been cloned in Dr. Patrick Sung's laboratory at Yale University (personal communications with Dr. Patrick Sung).

4.6 Detection interactions between BLM and TRF2

FRET (Fluorescence Resonance Energy Transfer) and co-immuoprecipitation results suggested that TRF2 and BLM interact with each other in ALT cells [52]. Purified TRF2 and BLM protein (BLM full-length protein was expressed and purified from budding

yeast *S. cerevisiae*) also could bind *in vitro* with high affinity (apparent *K*d = 2.5 nM) as suggested by ELISA assay results [61]. Therefore, it was believed that BLM is involved in telomere maintenance and processing. To further characterize the direct interaction between BLM and TRF proteins, I performed a direct binding assay between purified BLM-core and TRF1 and TRF2 using EMSA. As shown in Fig. 4.9, incubation of BLM-core with TRF2 resulted in a slower migration band, compared with BLM-core or TRF2 alone, suggesting a direct interaction between BLM-core and TRF2. In contrast, no difference was observed between BLM-core and TRF1 (Fig. 4.9). Thus, BLM-core only mediates the interaction with TRF2. Further mapping will be performed to identify the BLM-core binding domain of TRF2 before I can carry out structural study to establish the connection between the BTR complex and telomeres at atomic resolution by X-ray crystallography.

4.7 Progress towards the structure of the BTR complex

Human BLM protein contains three domains that are conserved throughout the RecQ helicase family. From the N- to the C-termini, these domains are the helicase domain, the RecQ carboxy-terminal (RQC) domain, and the Helicase and RNase D C-terminal (HRDC) domain [62] (Fig. 4.11). The helicase domain is essential for ATP binding and hydrolysis and defines the RecQ family. The RQC domain is also restricted to RecQ family members and is considered important for both the structural integrity of the protein and dsDNA binding. The HDRC domain has an auxiliary role in nucleic acid binding. These domains form the catalytic core of BLM (hereafter referred to as BLM-core, residues 646-1292). BLM-core was cloned into the His6-SUMO vector for protein

expression. However, initial purification of BLM-core yielded soluble aggregate. Slightly increase of the salt concentration from 150 to 250 mM NaCl helped prevent the aggregation. BLM-core with high purity then was subjected to intensive crystallization screening trial. Small needle crystals started to show up after 10 days of incubation at 4 °C (Fig. 4.10). Additional screening and optimization BLM-core will be performed to obtain crystals suitable for diffraction analysis.

The N-terminal half of BLM (BLM-N, residues 1-645) mediates the interactions with TOPOIIIα [42, 63]. However, primary sequence analysis indicated that BLM-N is mostly unstructured. Probably not surprisingly, expression of BLM-N alone failed because of low yield and insolubility. To address this issue, I am using both biochemical and yeast two-hybrid methods to map the interactions between BLM-N and TOPOIIIα. Similarly, the BLM-RMI1 and the TOPOIIIα-RMI1 interactions are also studied by the same approaches. (More detailed protein-protein interactions of BTR complex is shown in Fig. 4.11) The resulting information will be used to reconstitute the BTR complex *in vitro* for the structural study.

4.8 Methods and materials

Protein expression and purification

The C-terminal domain of RMI1 (residues 475 – 625) was cloned into a GST fusion protein expression vector, pGEX6p-1 (GE Healthcare). RMI1 (residues 6 – 147) was cloned into a modified pET28b vector with a SUMO protein fused at the amino-terminus after the His_6-tag [64]. The RMI1C-RMI2 complex was expressed in *E. coli.* Rosetta DE3 strain. SUMO-tagged RMI1C-OB and GST-tagged RMI2-OB plasmids were co-

transformed into Rosetta DE3 strain and the colonies carrying two plasmids were picked by Kana/Amp double selection. After induction for 16 hours with 0.1 mM IPTG at 25 °C, the cells were harvested by centrifugation and the pellets were resuspended in lysis buffer (50 mM Tris-HCl pH 8.0, 50 mM NaH_2PO_4, 400 mM NaCl, 3 mM imidazole, 10% glycerol, 1mM PMSF, 0.1 mg/ml lysozyme, 2mM 2-mercaptoethanol, and home-made protease inhibitor cocktail). The cells were then lysed by sonication and the cell debris was removed by ultracentrifugation. The supernatant was mixed with glutathione sepharose beads (GE Healthcare) and rocked for 6 hours at 4 °C before elution with 15 mM reduced glutathione. The protease 3C was added to remove the GST-tag. The complex was then mixed with Ni-NTA agarose beads (Qiagen) and rocked for 8 hours at 4 °C before elution with 250 mM imidazole. The ULP1 protease was added to remove the His_6-tag. Finally the RMI1C-RMI2 complex was further purified by passage through Mono-Q ion-exchange column and by gel-filtration chromatography on Hiload Superdex 75 equilibrated with 25 mM Tris-HCl pH 8.0, 150 mM NaCl and 5 mM dithiothreitol (DTT). The purified RMI1C-RMI2 complex protein was concertrated to 25 mg/ml and stored at -80 °C. The Seleno-Met substituted RMI1C-RMI2 complex protein was similarly purified.

The protein of N-terminal domain of RMI1 and BLM-core (rediues 646 – 1292) were expressed in *E. coli.* and purified following the same procedures as described above except for only one affinity chromatography step (Ni-NTA agarose) was used for RMI1N.

The protein of hTOPOIIIα-N was expressed in the insect cells (Sf9 and Hifive) expression system. The protein was purified following the same procedures as described

above except for only one affinity chromatography step (glutathione sepharose beads) was used for hTOPOIIIα-N.

Crystallization, data collection and structure determination

RMIN: The RMI1N protein was crystallized by sitting-drop-vapor-diffusion at 4 °C. The precipitant/well solution contained 100 mM Tris-HCl pH 8.5, 18% PEG 3350, 300 mM NaSCN, 10 mM $NiCl_2$ and 10 mM DTT. Crystals were gradually transferred to a harvesting solution containing 100 mM Tris-HCl pH 8.5, 25% PEG 3350, 25% glycerol, 300 mM NaSCN, 10 mM $NiCl_2$ and 10 mM DTT before flash-frozen in liquid nitrogen for storage and data collection under cryogenic conditions (100 K). Se-Met-SAD (at Se peak wavelength) dataset with the resolution of 2.0 Å was collected at beam line 21ID-D at APS and processed using HKL2000 (Otwinowski and Minor, 1997). Crystals belong to space group $P3_121$ with one complex per asymmetric unit. Five selenium sites were located and refined, and SAD phases calculated using SHARP (E. d. La Fortelle, G. Bricogne, 1997). A model was automatically built into the modified experimental electron density using ARP/WARP (V. S. Lamzin, A. Perrakis, K. S. Wilson, 2001); the model was then further refined using simulated-annealing and positional refinement in CNS [65] with manual rebuilding using program O [66].

RMI1C-RMI2 complex: The RMI1C-RMI2 complex protein was crystallized by hanging-drop-vapor-diffusion at 4 °C. The precipitant/well solution contained 18% PEG 3350, 300 mM NaSCN, and 10 mM DTT. Crystals were gradually transferred to a harvesting solution containing 25% PEG 3350, 25% glycerol, 300 mM NaSCN, and 10

mM DTT before flash-frozen in liquid nitrogen for storage and data collection under cryogenic conditions (100 K). Se-Met-SAD (at Se peak wavelength) dataset with the resolution of 2.0 Å was collected at beam line 21ID-D at APS and processed using HKL2000 (Otwinowski and Minor, 1997). Crystals belong to space group $P2_12_12_1$ with one complex per asymmetric unit. Eleven selenium sites were located and refined, and SAD phases calculated using SHARP (E. d. La Fortelle, G. Bricogne, 1997). A model was automatically built into the modified experimental electron density using ARP/WARP (V. S. Lamzin, A. Perrakis, K. S. Wilson, 2001); the model was then further refined using simulated-annealing and positional refinement in CNS [65] with manual rebuilding using program O [66].

BLM-core: The BLM-core protein was crystallized by sitting-drop-vapor-diffusion at 4 °C. The precipitant/well solution contained 50mM HEPES pH 7.0, 1.6 M Li_2SO_4, 50 mM $MgCl_2$ and 10 mM DTT.

Pull down assays:

For pull-down assays, $(His)_6$-BLM (5 μg), $(His)_6$-TOPOIIIα (5 μg), RMI1N (5 μg), or $RMI1N^{\Delta Loop}$ (5 μg) were used. The indicated proteins were incubated in 30 μl of Buffer B (25 mM Tris-HCl, pH 7.5, 0.005% Triton, 80 mM KCl, and 1 mM DTT) for 30 min at 4°C. The reactions were mixed with 11 μl of Ni-NTA agarose beads (which recognize the His-tag at the N-terminus of BLM and TOPOIIIα) at 4°C for 30 min. After washing the beads twice with 200 μl of the same buffer, bound proteins were eluted with 25 μl of 2%

SDS. The supernatant (S), wash (W), and SDS eluate (E), 10 μl each, were analyzed by 15% SDS-PAGE and Coomassie Blue staining.

dHJ dissolution assay:

The dHJ dissolution reaction is carried out in the presence of BLM (10 nM), TOPOIIIα (125 nM), and RMI1 (MBP-RMI1, RMI1N, RMI1$^{\Delta Loop}$) (140, 280 and 420nM) were incubated for 10 min on ice in 12 μl of reaction buffer (50 mM Tris-HCl, pH 7.8, 1 mM DTT, 0.8 mM MgCl2, 200 μg/ml bovine serum albumin, 2 mM ATP, 80 mM of KCl, and an ATP regenerating system consisting of 10 mM creatine phosphate and 50 μg/ml creatine kinase) followed by the addition of the dHJ substrate (1nM). After a 5-min incubation at 37 °C, 2 μl of 1% SDS and 1 μl of proteinase K (10 μg/μl stock) were added to the reaction mixtures, followed by a 10 min incubation at 37 °C. The deproteinized reactions were mixed with an equal volume of sample loading buffer (20 mM Tris-HCl, pH 7.5, 50% glycerol, and 0.08% Orange G) containing 50% urea, incubated at 95 °C for 3 min, and then resolved in 8% denaturing gels.

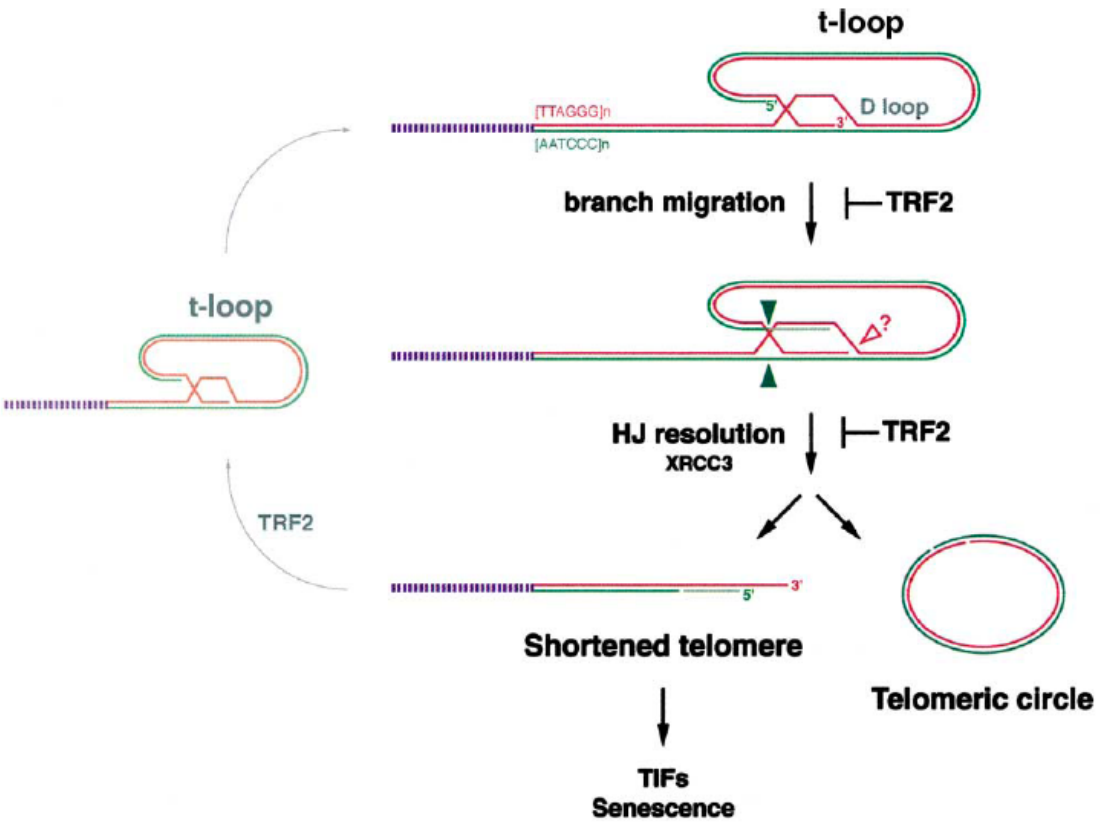


Figure 4.1* Speculative Model for T-loop HR
(Top) Proposed structure of t-loops. (Middle) Branch migration at the strand invasion site of the telomere terminus results in the formation of a Holliday junction. (Bottom) Two steps lead to t-loop deletion. Cleavage of the C strand at two positions by HJ resolvase (green arrowheads). This process is proposed to require XRCC3. The second step involves nicking of the D loop by an unknown nuclease (open arrowhead). The products are a shortened telomere and a relaxed telomeric circle. The shortened telomere might reform a small t-loop in a TRF2-dependent manner (left) or, if the deletion is too extensive, might activate a DNA damage response and induce senescence. TRF2 is proposed to promoted t-loop formation, thereby crating the substrate for t-loop HR. The basic domain of TRF2 is proposed to suppress t-loop HR by inhibiting branch migration and/or strand cleavage. In ALT cells, the telomeric circles resulting from t-loop HR could function as a template for rolling circle replication and allow telomeric-independent telomere maintenance.
(* This figure is photocopied from ref. [25], Fig. 7)

Figure 4.2* Proposed roles of the BTR complex in processing HR intermediates. Diagrammatic representation of the putative roles of the human BTR (BLM-TOPOIIIα-RMI1-RMI2) complex in processing HR intermediates during S-phase. Two homologous chromosomes are shown, labelled red and black. In this example, the initiating lesion is a ssDNA gap (on the red chromosome), such as is predicted to arise during discontinuous DNA synthesis (e.g. following bypass of a lesion on the lagging strand template).
(a) RAD51 catalyses the early strand invasion step of HR repair of ssDNA gaps to create a D-loop intermediate. At this stage, the human RecQ helicase, BLM, could prevent HR from progressing any further by disrupting D-loops.
(b) If homologous recombination repair does proceed, a key HR intermediate known as the 'double Holliday junction' (dHJ) forms. These dHJ intermediates are processed by either the BLM-TOPOIIIα-RMI1-RMI2 complex or putative HJ resolvases. If the latter occurs, then, depending on the relative orientation of HJ resolvase cleavage and re-joining of DNA fragments, crossing over of genetic material flanking the original repair site might occur. Alternatively, the BLM-TOPOIIIα-RMI1-RMI2 complex catalyses a reaction known as 'dHJ dissolution', which acts to resolve dHJs without any crossing over of genetic material flanking the repair site. This reaction encompasses two key steps: **(c)** convergent branch migration of individual HJs to create a hemicatenane intermediate, and **(d)** the decatenation of this structure. RMI1 stimulates TOPOIIIα activity, and probably performs a DNA-targeting function *in vivo*. Note, for simplicity the figure only shows the human proteins and mechanism, where BLM is the RecQ helicase homologue. Therefore the BTR complex is BLM-TOPOIIIα-RMI1-RMI2 in this figure.
(* This figure is photocopied from ref. [47] Fig. 1)

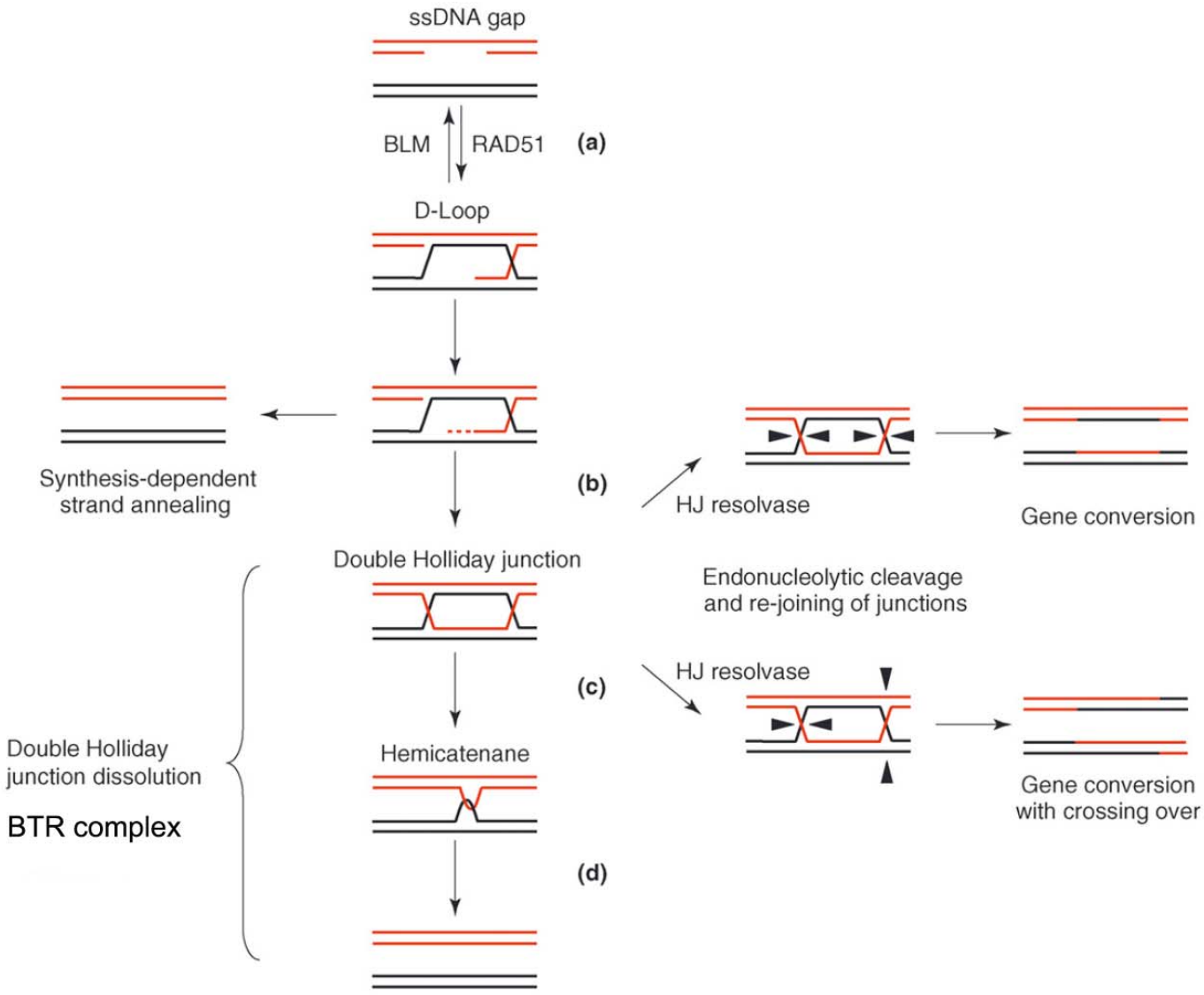
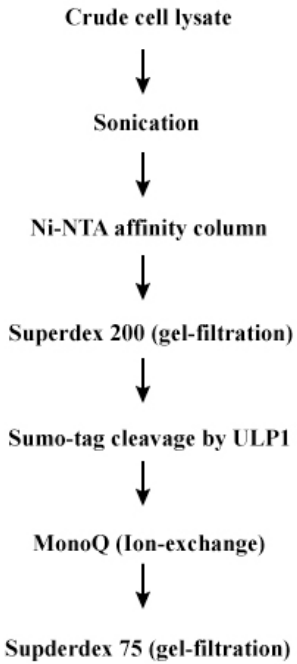
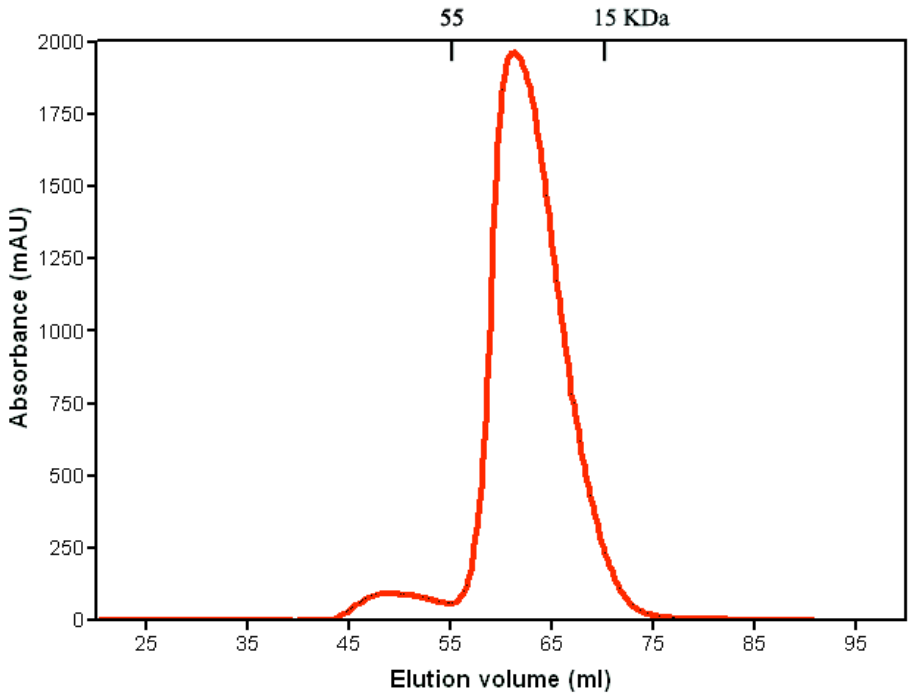

ssDNA gap
BLM
RAD51
(a)
D-Loop
Synthesis-dependent
strand annealing
(b)
HJ resolvase
Gene conversion
Double Holliday junction
Endonucleolytic cleavage
and re-joining of junctions
HJ resolvase
(c)
Double Holliday
junction dissolution
BTR complex
Hemicatenane
(d)
Gene conversion
with crossing over

Figure 4.3 Protein purification and crystallization of RMI1N
(a) Schematic diagram of purification procedure of RMI1N.
(b) Gel filtration chromatography profile (Hiload Superdex 75) of RMI1N protein. Elution positions of 15 and 55 KDa protein markers are indicated.
(c) SDS-PAGE of purified protein of RMI1N for crystallization.
(d) Crystal picture of RMI1N.

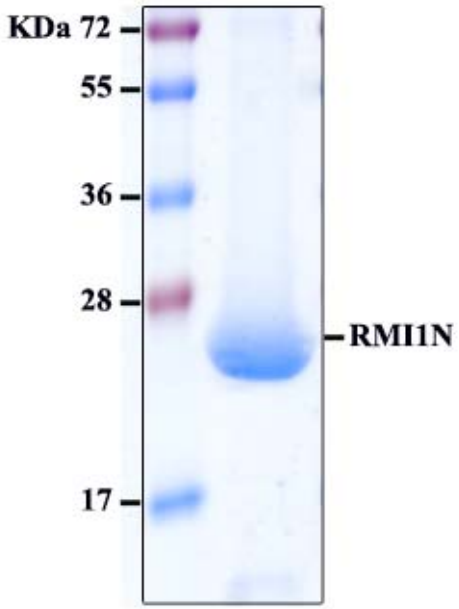
a


Crude cell lysate
Sonication
Ni-NTA affinity column
Superdex 200 (gel-filtration)
Sumo-tag cleavage by ULP1
MonoQ (Ion-exchange)
Supderdex 75 (gel-filtration)

b


55
15 KDa
2000
1750
1500
1250
1000
750
500
250
0
Absorbance (mAU)
25
35
45
55
65
75
85
95
Elution volume (ml)

c



d

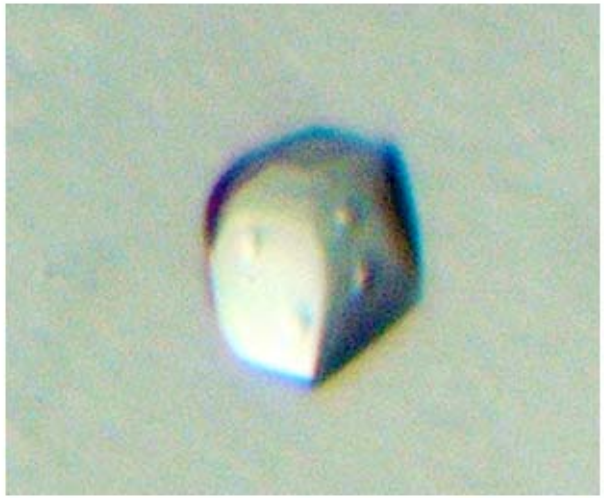
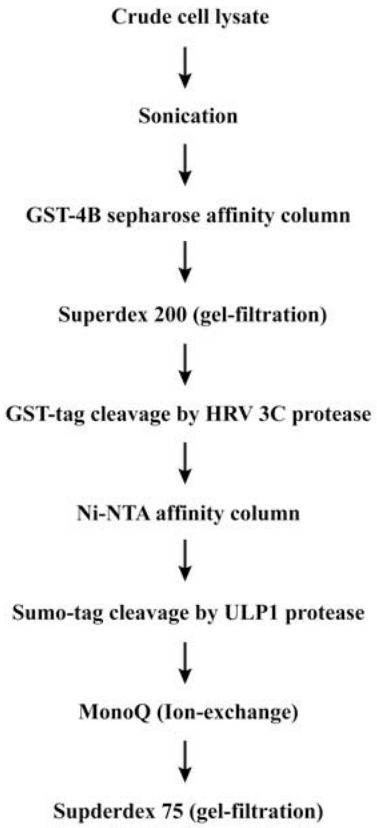


Table 4.1 Data collection, phasing and refinement statistics for RMI1N

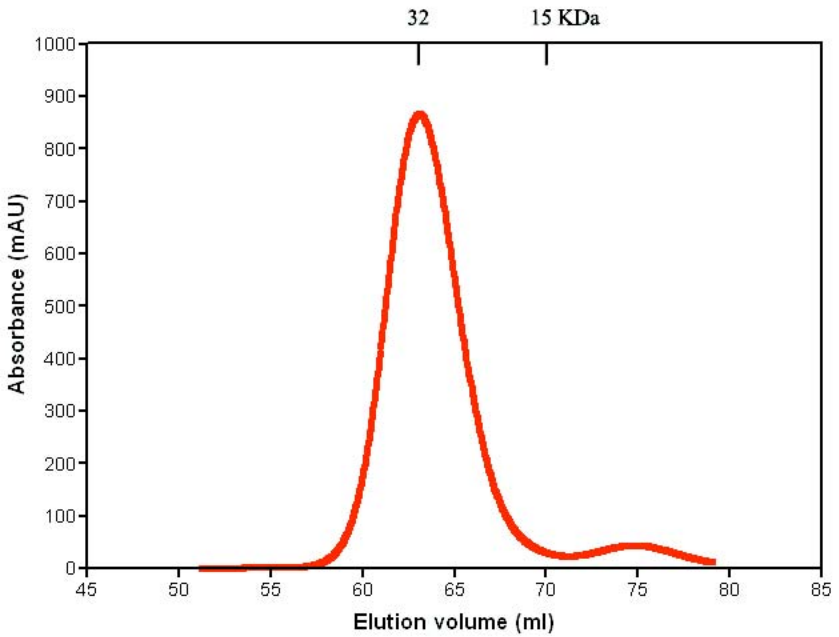
	RMI1N
Data collection	
	Se Peak
Space group	$P3_121$
Cell dimensions	
a, *b*, *c* (Å)	57.896, 57.896, 117.475
α, β, γ (º)	90, 90, 120
Wavelength (Å)	0.97872
Resolution (Å)	100-2.0
R_{merge} (%) (high res. shell)	0.073 (0.409)
I/σ (high res. cell)	48.9 (4.5)
Completeness (%) (high res. shell)	99.1 (94.7)
Redundancy (high res. shell)	14.0 (11.3)
Phasing	
Acentric Phasing Power	1.734
Acentric reflections FOM	0.385
Centric reflections FOM	0.114
Refinement	
Resolution (Å)	50-2.0
No. reflections	34177
R_{work}/ R_{free} (%)	22.2/25.6
B-factors (Å^2)	
Protein	42.8
Water	41.0
R.m.s deviations	
Bond lengths (Å)	0.005823
Bond angles (º)	1.43358

Figure 4.4 Protein purification and crystallization of RMI1C-RMI2 complex
(a) Schematic diagram of purification procedure of RMI1C-RMI2 complex.
(b) Gel filtration chromatography profile (Hiload Superdex 75) of RMI1C-RMI2 complex. Elution positions of 15 and 32 KDa protein markers are indicated.
(c) SDS-PAGE of purified protein of RMI1C-RMI2 complex for crystallization.
(d) Crystal picture of RMI1C-RMI2 complex.

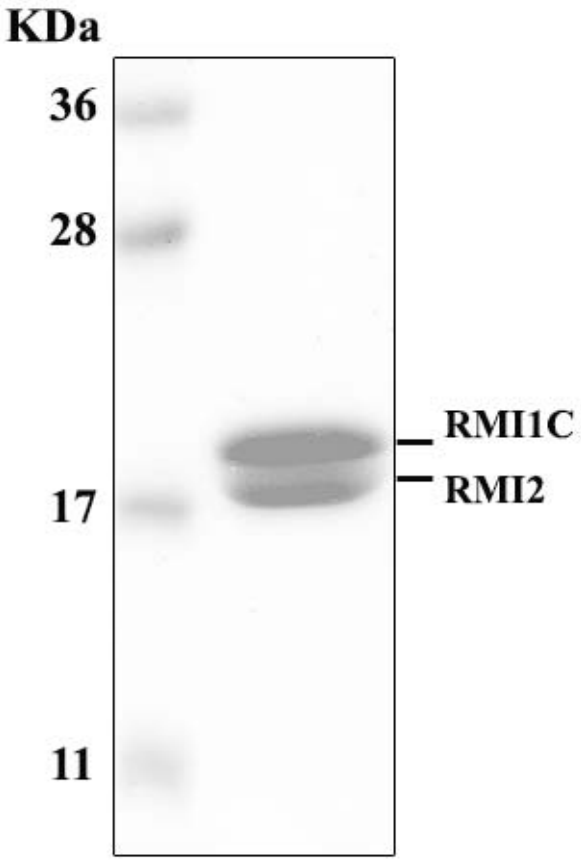
a


Crude cell lysate
Sonication
GST-4B sepharose affinity column
Superdex 200 (gel-filtration)
GST-tag cleavage by HRV 3C protease
Ni-NTA affinity column
Sumo-tag cleavage by ULP1 protease
MonoQ (Ion-exchange)
Supderdex 75 (gel-filtration)

b


32
15 KDa
Absorbance (mAU)
0
100
200
300
400
500
600
700
800
900
1000
45
50
55
60
65
70
75
80
85
Elution volume (ml)

c



d

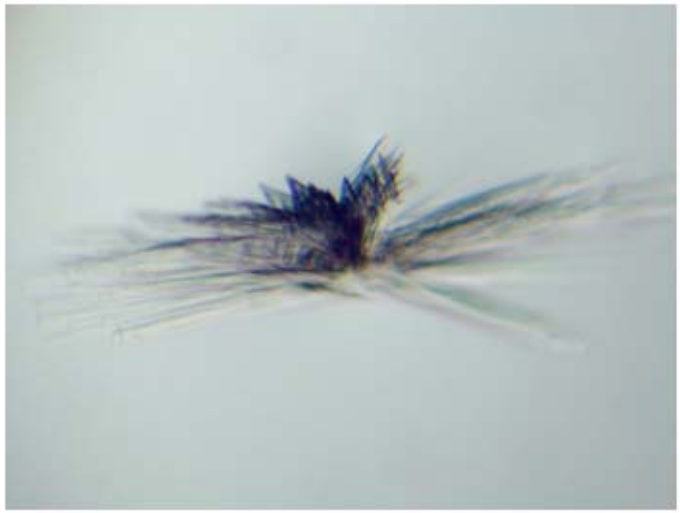
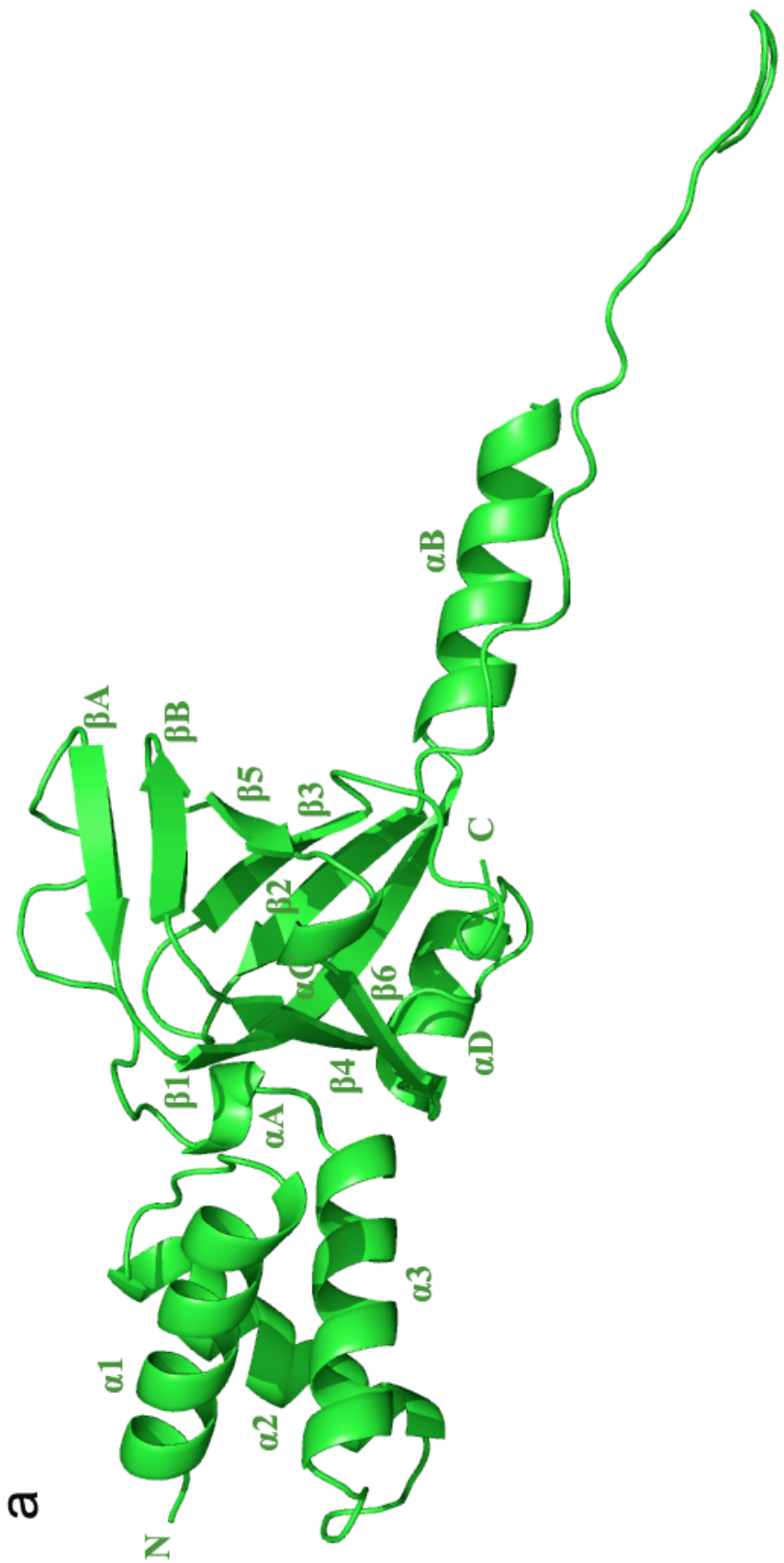
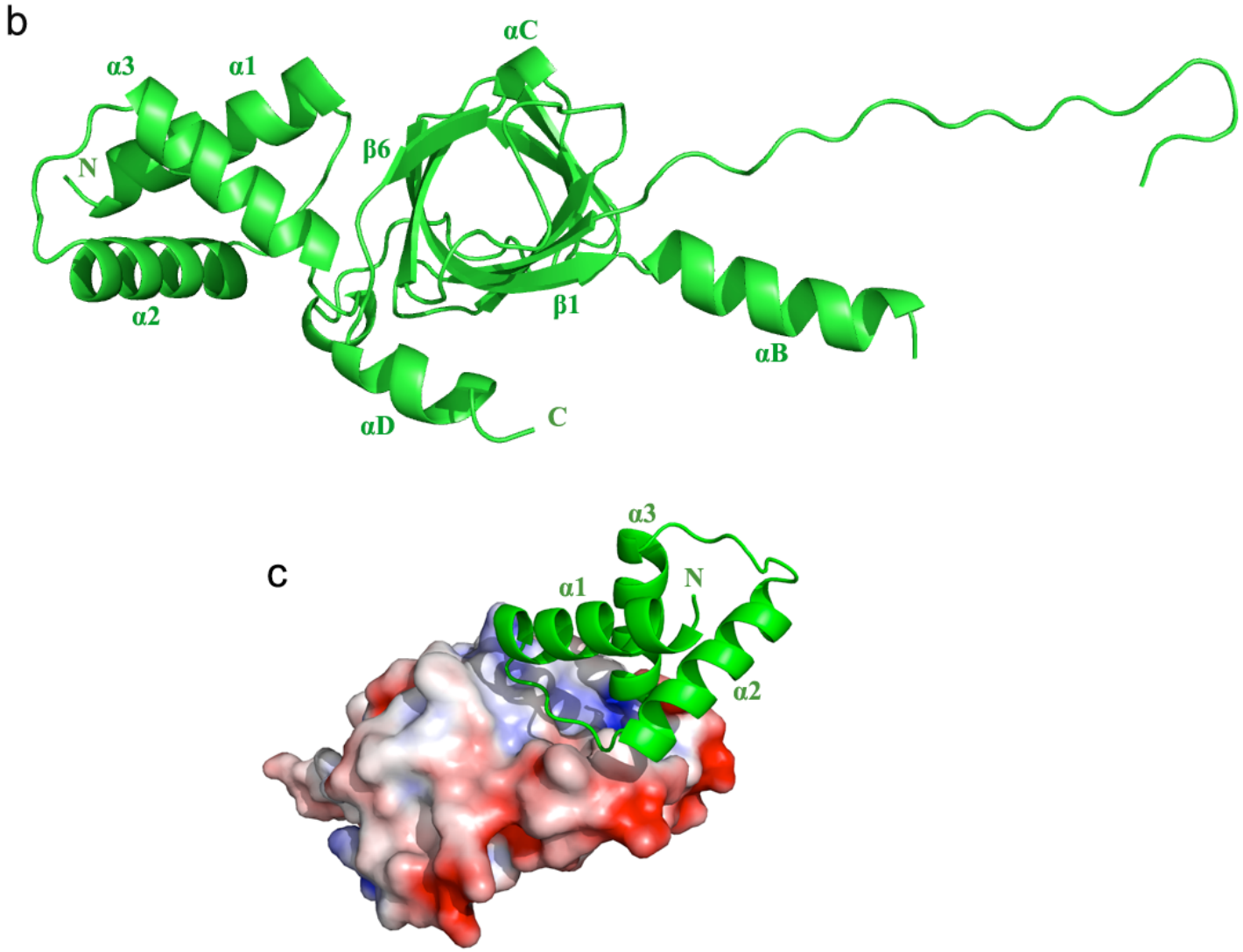


Table 4.2 Data collection, phasing and refinement statistics for RMI1C-RMI2

	RMI1C-RMI2
Data collection	
	Se Peak
Space group	$P2_12_12_1$
Cell dimensions	
a, *b*, *c* (Å)	41.682, 42.365, 157.734
α, β, γ (º)	90, 90, 120
Wavelength (Å)	0.97934
Resolution (Å)	100-2.0
R_{merge} (%) (high res. shell)	0.112 (0.279)
I/σ (high res. cell)	33.1 (6.4)
Completeness (%) (high res. shell)	98.4 (89.7)
Redundancy (high res. shell)	12.0 (7.1)
Phasing	
Acentric Phasing Power	1.990
Acentric reflections FOM	0.477
Centric reflections FOM	0.139
Refinement	
Resolution (Å)	50-2.0
No. reflections	35460
R_{work}/ R_{free} (%)	23.8/25.8
B-factors (Å^2)	
Protein	14.584
Water	30.612
R.m.s deviations	
Bond lengths (Å)	0.008171
Bond angles (º)	2.25049

Figure 4.5 Crystal structure of RMI1N
(a) RMI1N-OB adopts a typical OB-fold structure.
(b) Bottom view of RMI1N-OB structure. Neither side of the RMI1N OB-fold is covered by any α helices.
(c) The N-terminal three-helix-bundle motif of RMI1N packs on the OB-fold though hydrophobic van der Waals contacts. The OB-fold is in surface representation and colored according to its electrostatic potential (positive potential, blue; negative potential, red).
(d) Superposition of RMI1N-OB (green) on the crystal structure of human RPA1N-OB (yellow).


a
N
α1
α2
α3
αA
β1
βA
βB
β5
β3
β2
αC
β6
β4
αD
C
αB


b
αC
α3
α1
β6
N
α2
β1
αB
αD
C
c
α3
α1
N
α2

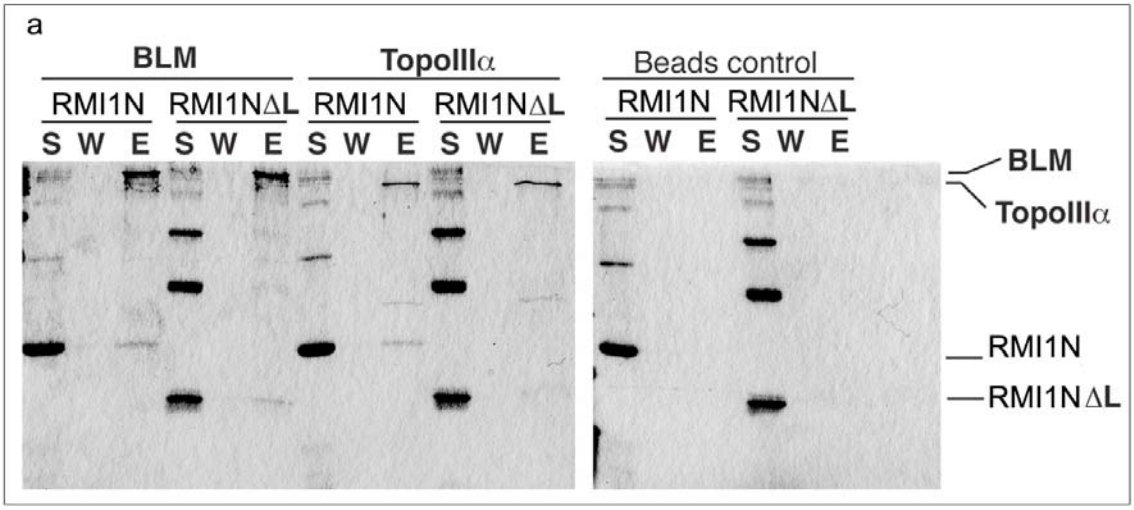
d

RMI1-OB / RPA1

Figure 4.6 *In vitro* tests of RMI1N and RMI1N$^{\Delta loop}$ (Done by Dr. Patrick Sung's group)

(a) RMI1N and RMI1N$^{\Delta loop}$ interact with both BLM and TOPOIIIα. RMI1N and RMI1N$^{\Delta loop}$ can be detected in elution (E) lanes after incubating with His6-BLM /His6-TOPOIIIα on Ni-NTA. RMI1N and RMI1N$^{\Delta loop}$ cannot be seen in supernatant (S), wash (W) and Ni-NTA control lanes.

(b) RMI1N enhances the dissolution activity of BLM- TOPOIIIα, while RMI1N$^{\Delta loop}$ doesn't. Increased amount of RMI1N and RMI1N$^{\Delta loop}$ proteins together with BLM-TOPOIIIα are incubated with radioactive labeled dHJ substrates (upper band). The dissolution products (lower band) are detected by Phosphorimager. The percentage of the dissolution reaction is quantified and charted blow.


a
BLM
TopoIIIα
Beads control
RMI1N
RMI1NΔL
RMI1N
RMI1NΔL
RMI1N
RMI1NΔL
S W E S W E S W E S W E S W E S W E
BLM
TopoIIIα
RMI1N
RMI1NΔL

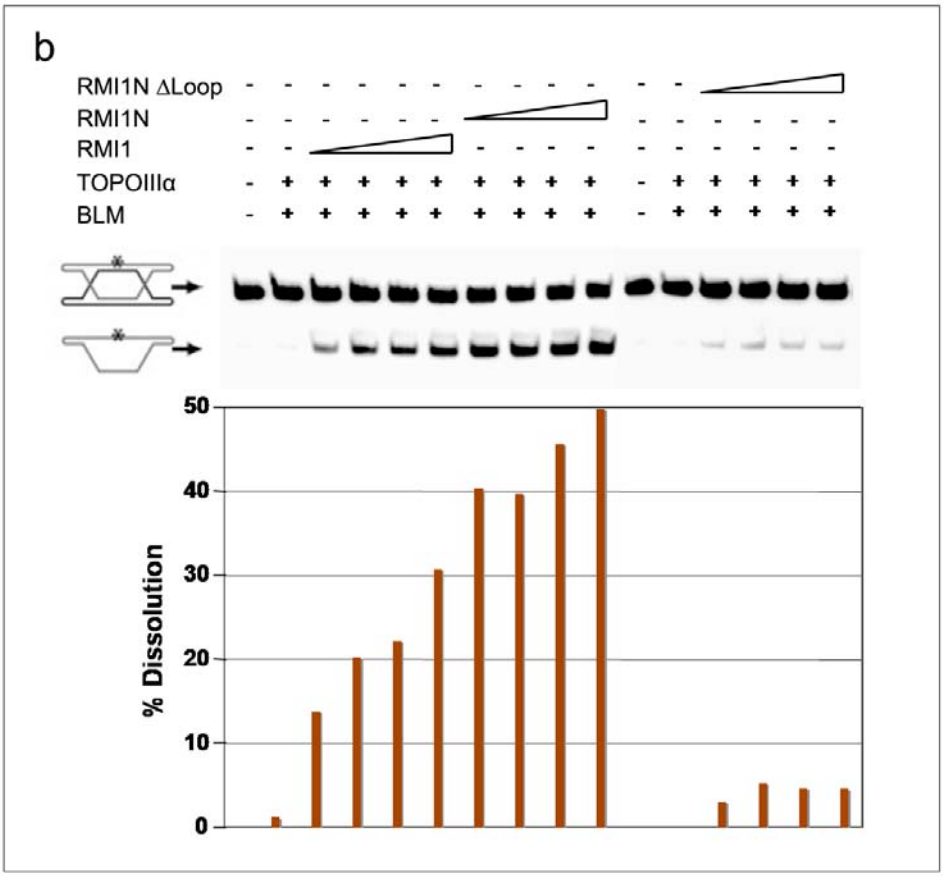
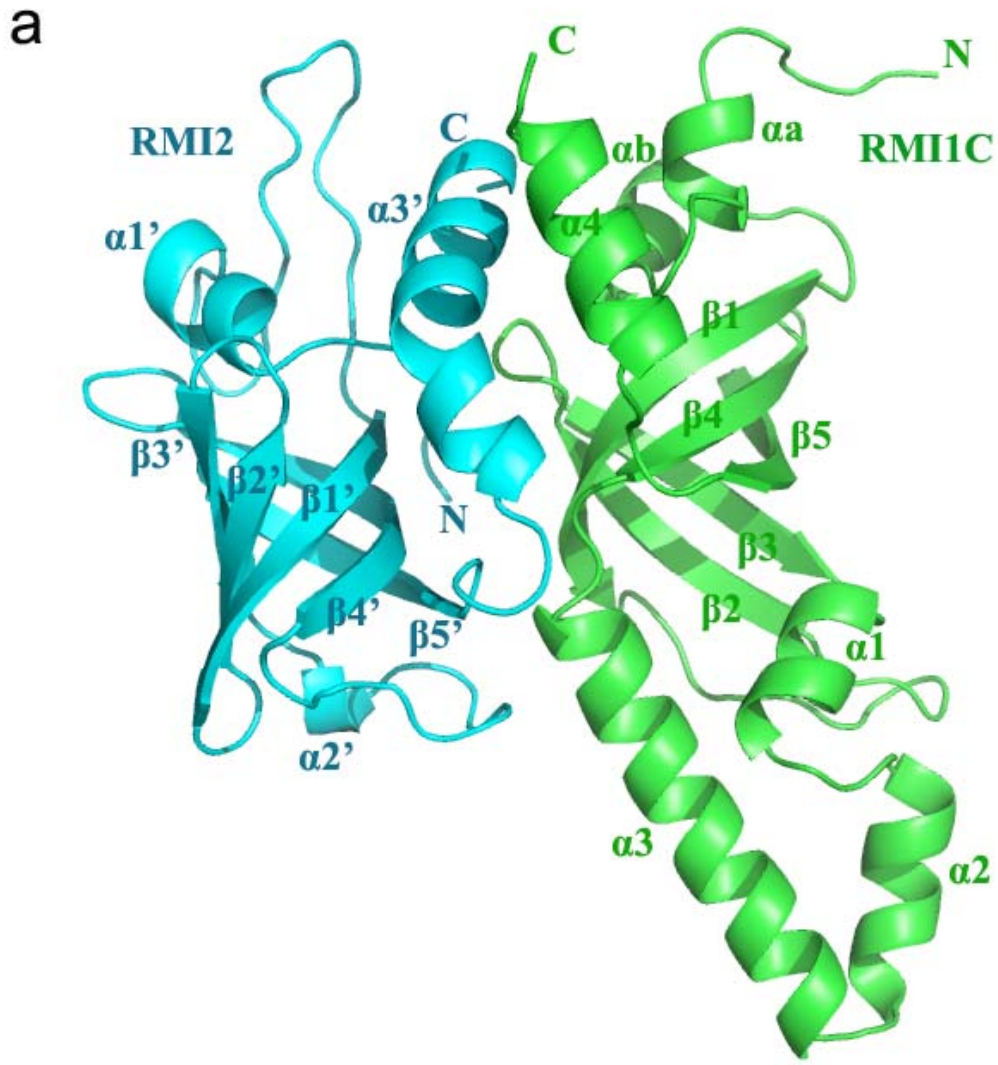
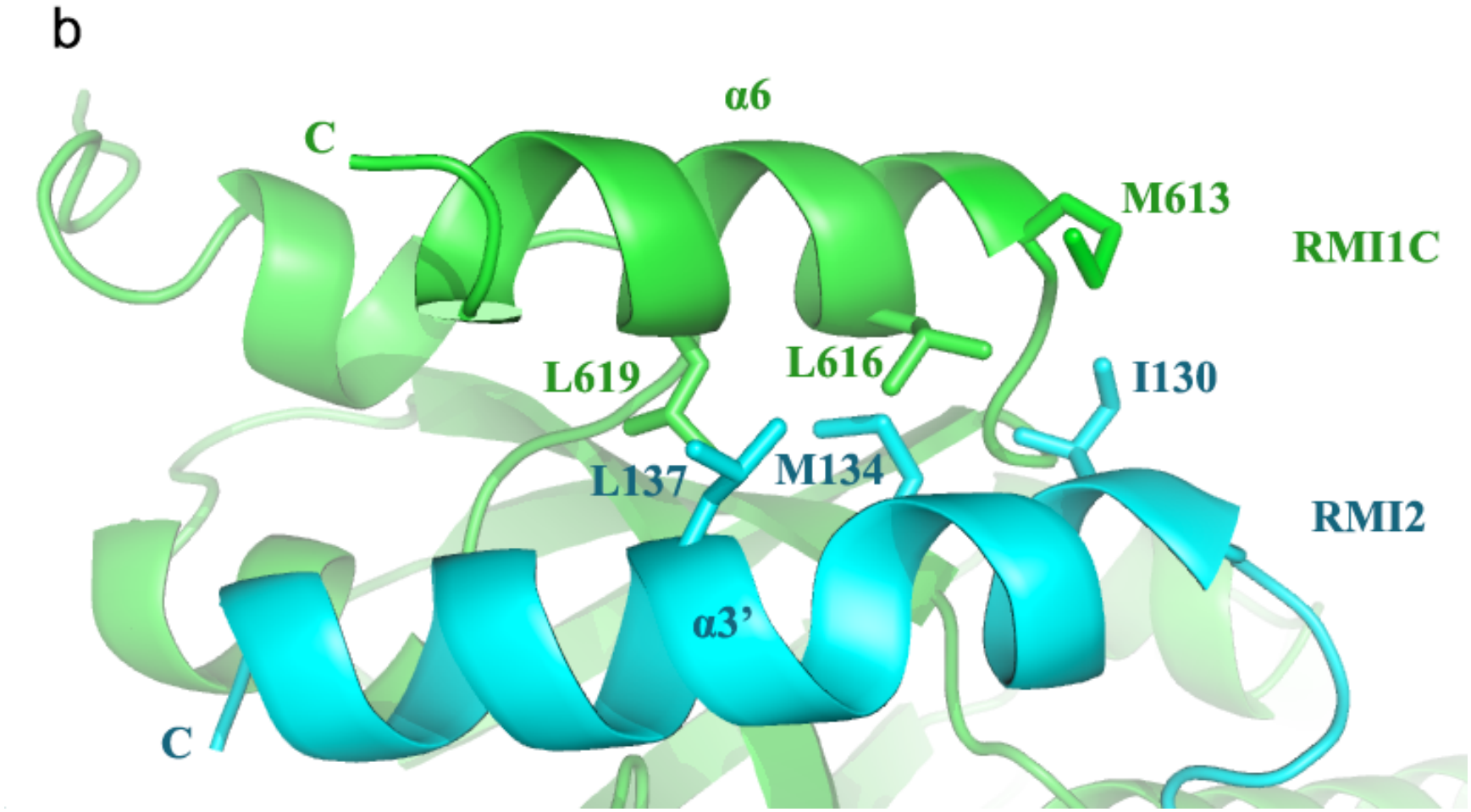

b
RMI1N ΔLoop
RMI1N
RMI1
TOPOIIIα
BLM
% Dissolution
0
10
20
30
40
50

Figure 4.7 Crystal structure of RMI1C-RMI2 complex
(a) Both RMI1C and RMI2 contain an OB-fold. The C-terminal α-helices from RMI1C and RMI2 medicate the protein interaction.
(b) Hydrophobic residues locating on each α helix from RMIC (Met613, Leu616 and Leu619) and RMI2 (Ile130, Met134 and Leu137) interdigitate with one another and form the major hydrophobic interface.
(c) Two salt-bridge interactions formed between Arg622 in RMI1C and Asp141 in RMI2.


a
RMI2
C
C
N
αb
αa
RMI1C
α3'
α1'
α4
β1
β4
β5
β3'
β2'
β1'
N
β3
β2
α1
β4'
β5'
α2'
α3
α2


b
α6
C
M613
RMI1C
L619
L616
I130
L137
M134
RMI2
α3'
C

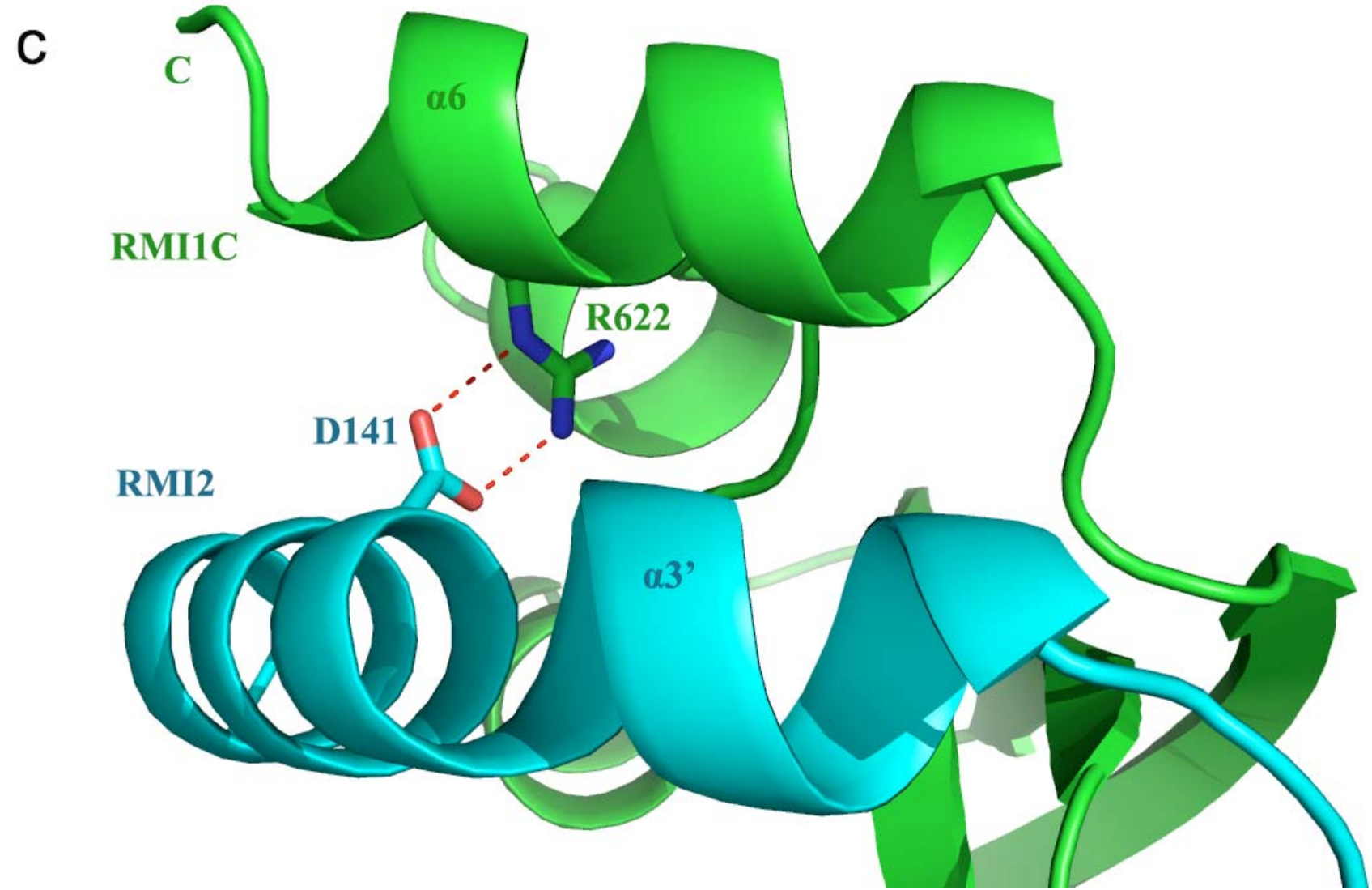
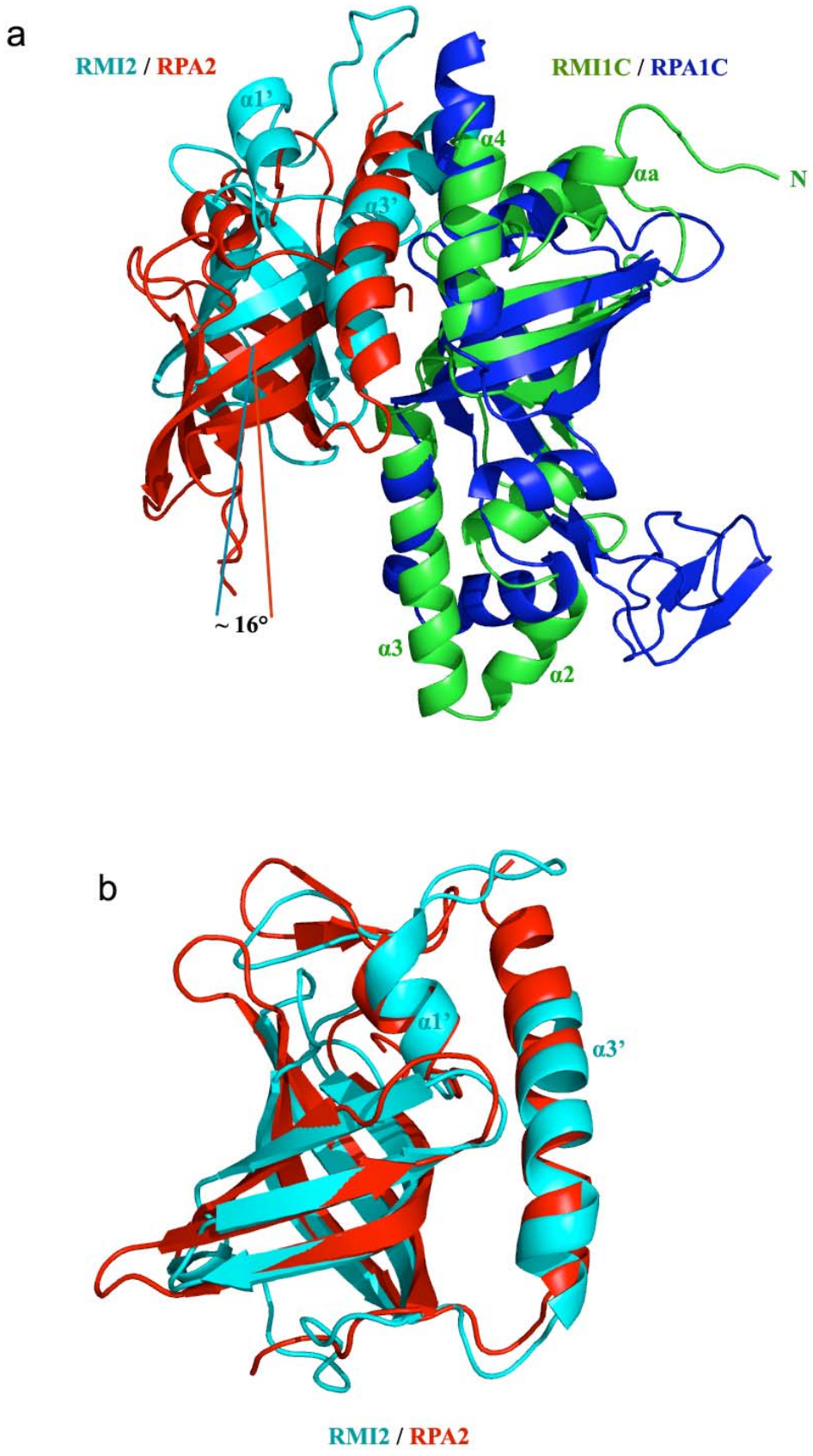
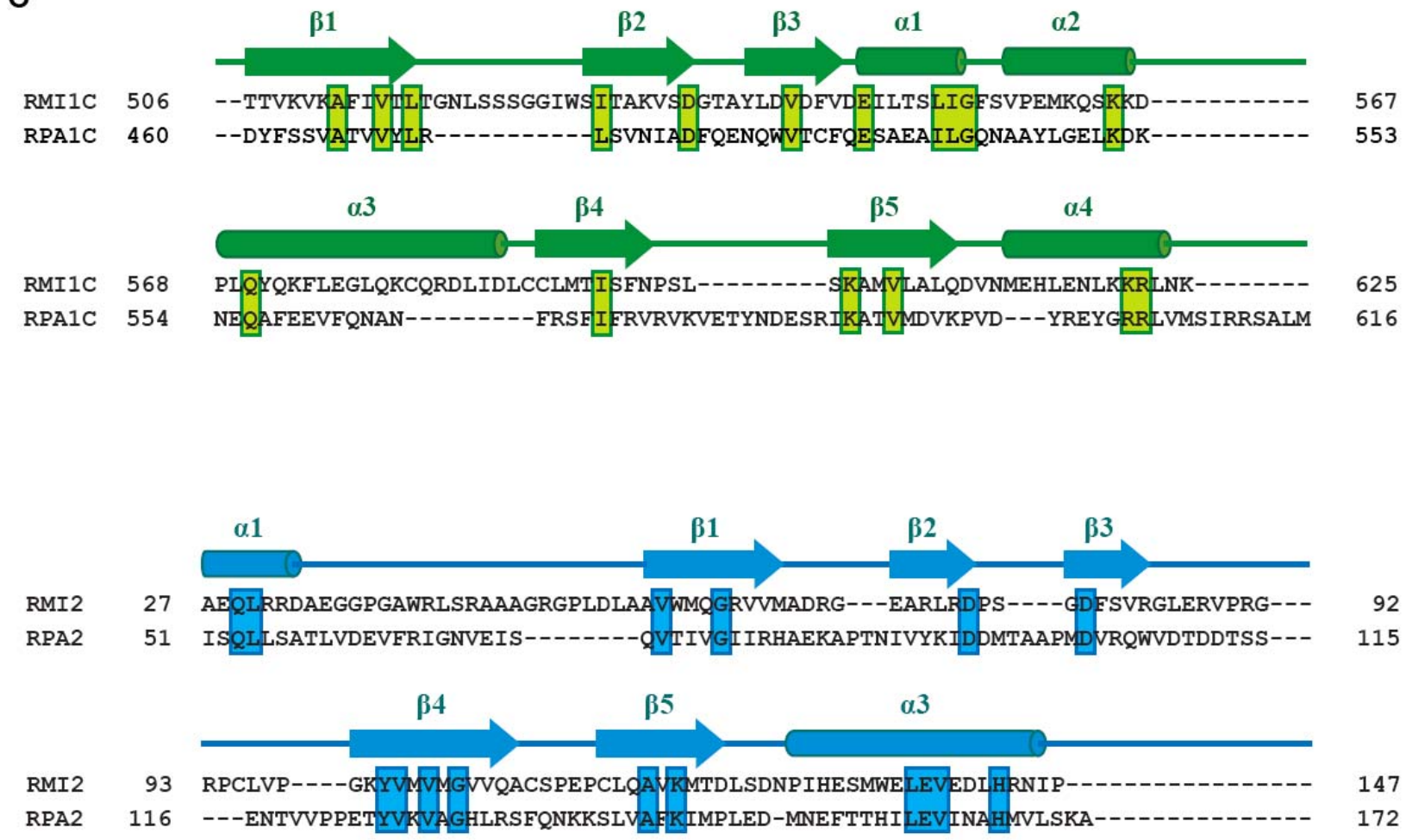

C
C
α6
RMI1C
R622
D141
RMI2
α3'

Figure 4.8 RMI1C-RMI2 complex is structural conserved with RPA1C-RPA2 complex.
(a) Superposition of RMI1C-RMI2 complex (RMI1C: green, RMI2: cyan) on the crystal structure of human RPA1C-RPA2 complex (RPA1C: blue, RPA2: red).
(b) Superposition of RMI2 (RMI2: cyan) on the crystal structure of human RPA2 (RPA2: red).
(c) Amino acid sequence alignment of RMI1C-RMI2 complex and human RPA1C-RPA2 complex*. Upper panel: sequence alignment of the C-terminal OB-fold regions of RMI1 and human RPA1. Lower panel: sequence alignment of OB-fold regions of RMI2 and human RPA2. The sequence alignments with RPA1C and RPA2 are based on the crystal structure of the human RPA trimerization core complex [67]. Secondary structural assignments from our RMI1C-RMI2 complex are shown as colored cylinders (α helices, green: RMI1C and RPA1) and arrows (β strands, cyan: RMI2 and RPA2) above the sequences. The highlighted areas indicate the conserved amino acids. (*The β strands (βa and βb) in RPA1C and α helix (α2) in RMI2 are not shown in the alignment since there are no such corresponding structures in RMI1C and RPA2.)


a
RMI2 / RPA2
RMI1C / RPA1C
α1'
α4
αa
N
α3'
~ 16°
α3
α2
b
α1'
α3'
RMI2 / RPA2

C


β1 β2 β3 α1 α2
RMI1C 506 --TTVKVKAFIVTLTGNLSSSGGIWSITAKVSDGTAYLDVDFVDEILTSLIGFSVPEMKQSKKD---------- 567
RPA1C 460 --DYFSSVATVVYLR----------LSVNIADFQENQWVTCFQESAEAILGQNAAYLGELKDK---------- 553
α3 β4 β5 α4
RMI1C 568 PLQYQKFLEGLQKCQRDLIDLCCLMTISFNPSL---------SKAMVLALQDVNMEHLENLKKRLNK-------- 625
RPA1C 554 NEQAFEEVFQNAN---------FRSFIFRVRVKVETYNDESRIKATVMDVKPVD---YREYGRRLVMSIRRSALM 616
α1 β1 β2 β3
RMI2 27 AEQLRRDAEGGPGAWRLSRAAAGRGPLDLAAVWMQGRVVMADRG---EARLRDPS----GDFSVRGLERVPRG--- 92
RPA2 51 ISQLLSATLVDEVFRIGNVEIS--------QVTIVGIIRHAEKAPTNIVYKIDDMTAAPMDVRQWVDTDDTSS--- 115
β4 β5 α3
RMI2 93 RPCLVP----GKYVMVMGVVQACSPEPCLQAVKMTDLSDNPIHESMWELEVEDLHRNIP---------------- 147
RPA2 116 ---ENTVVPPETYVKVAGHLRSFQNKKSLVAFKIMPLED-MNEFTTHILEVINAHMVLSKA--------------- 172

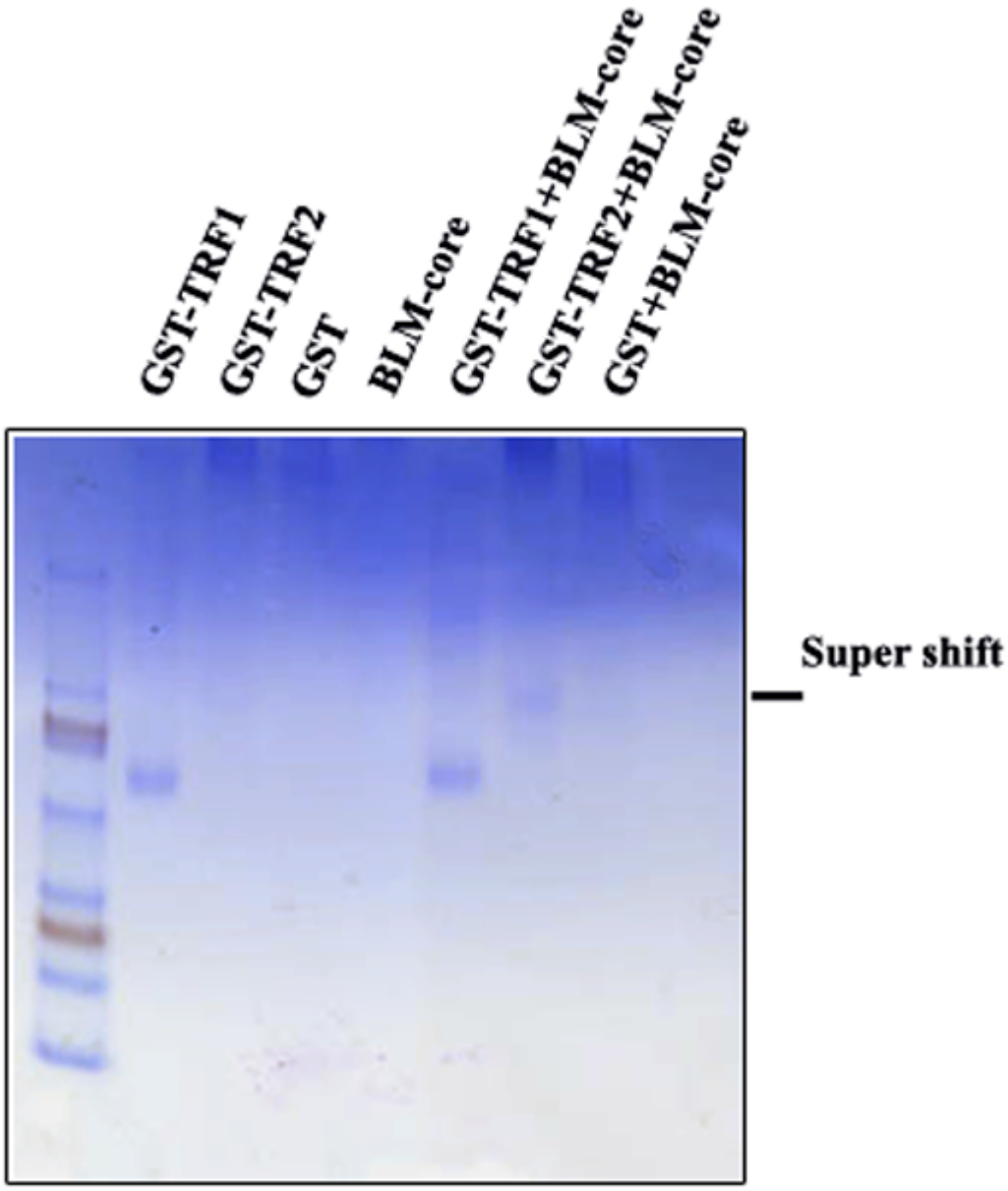


Figure 4.9 BLM interacts with TRF2 *in vitro*. Purified BLM-core protein is incubated with GST-TRF1 or GST-TRF2 protein for native gel. Only the incubation of BLM-core with TRF2 resulted in a slower migration pattern in the native gel while TRF1 didn't.

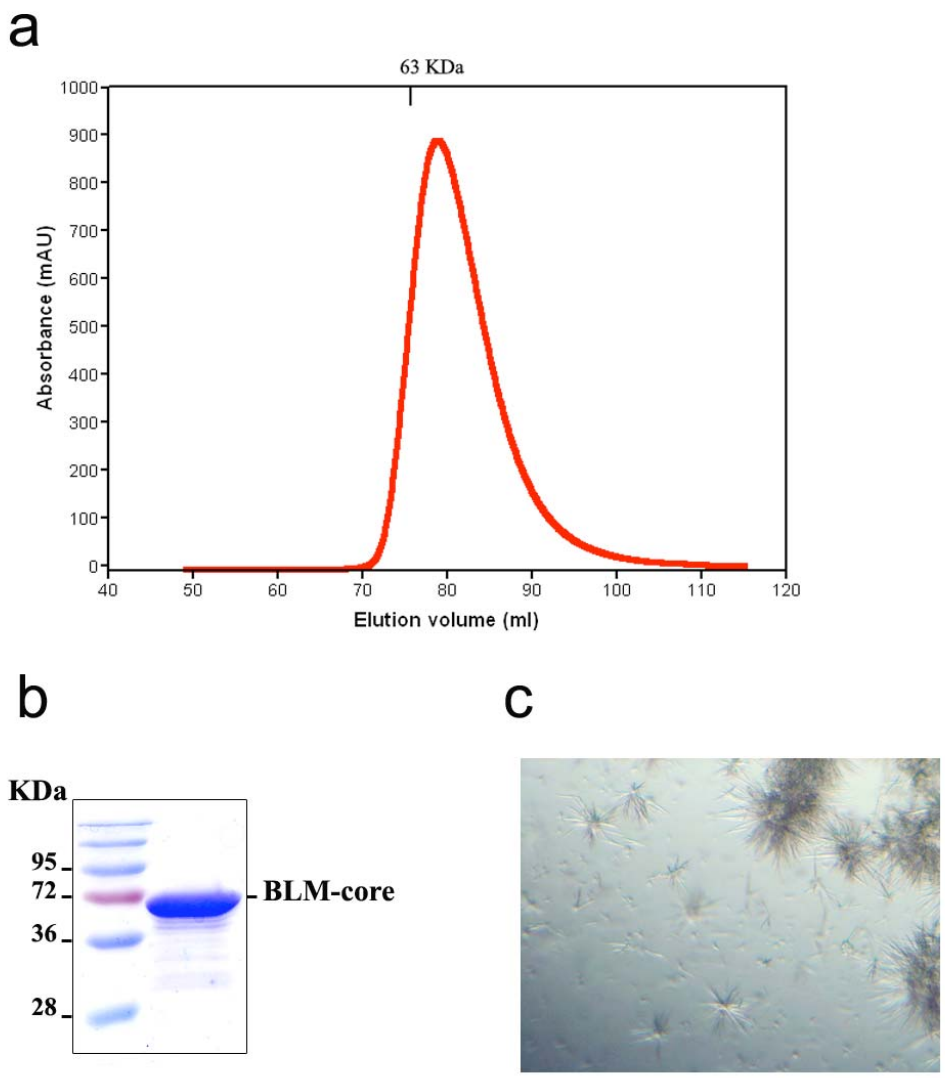
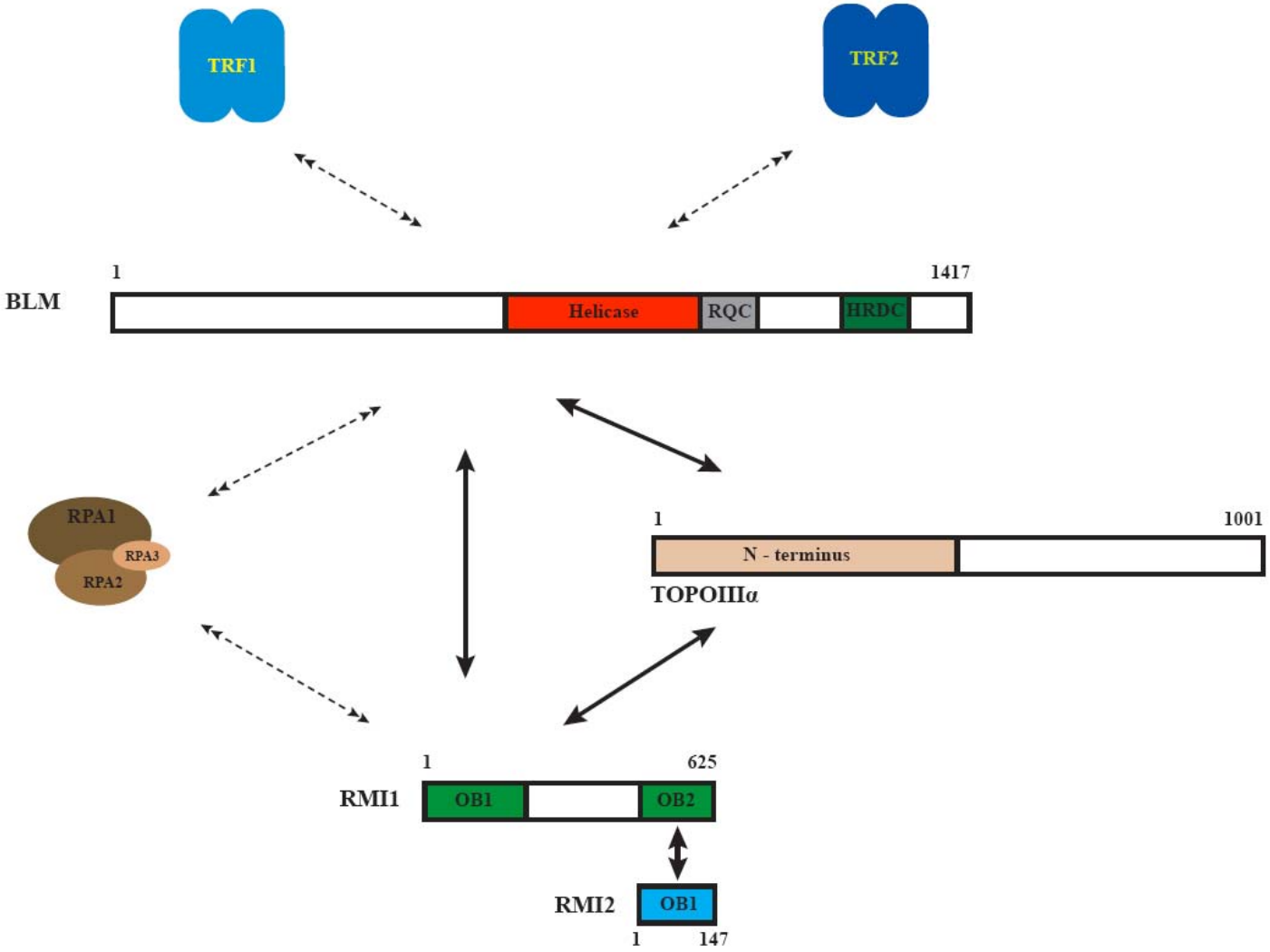


Figure 4.10 Protein purification and crystallization of BLM-core (646-1292)
(a) Gel filtration chromatography profile (Hiload Superdex 200) of BLM-core. Elution position of 63 KDa protein markers is indicated.
(b) SDS-PAGE of purified protein of BLM-core for crystallization.
(c) Crystal picture of BLM-core.

Figure 4.11 Schematic diagram of BLM-TOPOIIIα-RMI1/2 (BTR) complex and other potential interacting protein in human cells. Individual domain of each component of BTR complex is coloured and labeled. Direct interactions between two proteins are indicated by solid black double arrows; and the potential interactions are indicated by dashed black arrows.


TRF1
TRF2
BLM
1
1417
Helicase
RQC
HRDC
RPA1
RPA3
RPA2
1
1001
N - terminus
TOPOIIIα
RMI1
1
625
OB1
OB2
RMI2
OB1
1
147

Reference:

1. Kim, N.W., et al., *Specific association of human telomerase activity with immortal cells and cancer.* Science, 1994. **266**(5193): p. 2011-5.
2. Bryan, T.M., et al., *Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines.* Nat Med, 1997. **3**(11): p. 1271-4.
3. Hodges, M., et al., *Structure, organization, and dynamics of promyelocytic leukemia protein nuclear bodies.* Am J Hum Genet, 1998. **63**(2): p. 297-304.
4. Maul, G.G., et al., *Review: properties and assembly mechanisms of ND10, PML bodies, or PODs.* J Struct Biol, 2000. **129**(2-3): p. 278-87.
5. Ruggero, D., Z.G. Wang, and P.P. Pandolfi, *The puzzling multiple lives of PML and its role in the genesis of cancer.* Bioessays, 2000. **22**(9): p. 827-35.
6. Zhong, S., P. Salomoni, and P.P. Pandolfi, *The transcriptional role of PML and the nuclear body.* Nat Cell Biol, 2000. **2**(5): p. E85-90.
7. Borden, K.L., *Pondering the promyelocytic leukemia protein (PML) puzzle: possible functions for PML nuclear bodies.* Mol Cell Biol, 2002. **22**(15): p. 5259-69.
8. Dellaire, G. and D.P. Bazett-Jones, *PML nuclear bodies: dynamic sensors of DNA damage and cellular stress.* Bioessays, 2004. **26**(9): p. 963-77.
9. Yeager, T.R., et al., *Telomerase-negative immortalized human cells contain a novel type of promyelocytic leukemia (PML) body.* Cancer Res, 1999. **59**(17): p. 4175-9.
10. Murnane, J.P., et al., *Telomere dynamics in an immortal human cell line.* EMBO J, 1994. **13**(20): p. 4953-62.
11. Bryan, T.M., et al., *Telomere elongation in immortal human cells without detectable telomerase activity.* EMBO J, 1995. **14**(17): p. 4240-8.
12. Ogino, H., et al., *Release of telomeric DNA from chromosomes in immortal human cells lacking telomerase activity.* Biochem Biophys Res Commun, 1998. **248**(2): p. 223-7.
13. Regev, A., et al., *Telomeric repeats on small polydisperse circular DNA (spcDNA) and genomic instability.* Oncogene, 1998. **17**(26): p. 3455-61.
14. Tokutake, Y., et al., *Extra-chromosomal telomere repeat DNA in telomerase-negative immortalized cell lines.* Biochem Biophys Res Commun, 1998. **247**(3): p. 765-72.
15. Harley, C.B., A.B. Futcher, and C.W. Greider, *Telomeres shorten during ageing of human fibroblasts.* Nature, 1990. **345**(6274): p. 458-60.
16. Harley, C.B. and S.W. Sherwood, *Telomerase, checkpoints and cancer.* Cancer Surv, 1997. **29**: p. 263-84.
17. Wright, W.E., et al., *Normal human chromosomes have long G-rich telomeric overhangs at one end.* Genes Dev, 1997. **11**(21): p. 2801-9.
18. Martens, U.M., et al., *Accumulation of short telomeres in human fibroblasts prior to replicative senescence.* Exp Cell Res, 2000. **256**(1): p. 291-9.
19. Allshire, R.C., M. Dempster, and N.D. Hastie, *Human telomeres contain at least three types of G-rich repeat distributed non-randomly.* Nucleic Acids Res, 1989. **17**(12): p. 4611-27.

20. de Lange, T., et al., *Structure and variability of human chromosome ends.* Mol Cell Biol, 1990. **10**(2): p. 518-27.
21. Hastie, N.D., et al., *Telomere reduction in human colorectal carcinoma and with ageing.* Nature, 1990. **346**(6287): p. 866-8.
22. Counter, C.M., et al., *Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity.* EMBO J, 1992. **11**(5): p. 1921-9.
23. Grobelny, J.V., A.K. Godwin, and D. Broccoli, *ALT-associated PML bodies are present in viable cells and are enriched in cells in the G(2)/M phase of the cell cycle.* J Cell Sci, 2000. **113 Pt 24**: p. 4577-85.
24. Opitz, O.G., et al., *Cyclin D1 overexpression and p53 inactivation immortalize primary oral keratinocytes by a telomerase-independent mechanism.* J Clin Invest, 2001. **108**(5): p. 725-32.
25. Wang, R.C., A. Smogorzewska, and T. de Lange, *Homologous recombination generates T-loop-sized deletions at human telomeres.* Cell, 2004. **119**(3): p. 355-68.
26. Cesare, A.J. and J.D. Griffith, *Telomeric DNA in ALT cells is characterized by free telomeric circles and heterogeneous t-loops.* Mol Cell Biol, 2004. **24**(22): p. 9948-57.
27. Puranam, K.L. and P.J. Blackshear, *Cloning and characterization of RECQL, a potential human homologue of the Escherichia coli DNA helicase RecQ.* J Biol Chem, 1994. **269**(47): p. 29838-45.
28. Seki, M., et al., *Molecular cloning of cDNA encoding human DNA helicase Q1 which has homology to Escherichia coli Rec Q helicase and localization of the gene at chromosome 12p12.* Nucleic Acids Res, 1994. **22**(22): p. 4566-73.
29. Kitao, S., et al., *Cloning of two new human helicase genes of the RecQ family: biological significance of multiple species in higher eukaryotes.* Genomics, 1998. **54**(3): p. 443-52.
30. Yu, C.E., et al., *Positional cloning of the Werner's syndrome gene.* Science, 1996. **272**(5259): p. 258-62.
31. Lombard, D.B. and L. Guarente, *Cloning the gene for Werner syndrome: a disease with many symptoms of premature aging.* Trends Genet, 1996. **12**(8): p. 283-6.
32. Ellis, N.A., et al., *The Bloom's syndrome gene product is homologous to RecQ helicases.* Cell, 1995. **83**(4): p. 655-66.
33. German, J., et al., *Bloom syndrome: an analysis of consanguineous families assigns the locus mutated to chromosome band 15q26.1.* Proc Natl Acad Sci U S A, 1994. **91**(14): p. 6669-73.
34. McDaniel, L.D. and R.A. Schultz, *Elevated sister chromatid exchange phenotype of Bloom syndrome cells is complemented by human chromosome 15.* Proc Natl Acad Sci U S A, 1992. **89**(17): p. 7968-72.
35. Chester, N., et al., *Stage-specific apoptosis, developmental delay, and embryonic lethality in mice homozygous for a targeted disruption in the murine Bloom's syndrome gene.* Genes Dev, 1998. **12**(21): p. 3382-93.

36. Mori, I., et al., *Carcinogen dose-dependent variation in the transgene mutation spectrum in urethane-induced lung tumors in transgenic mice carrying the human prototype c-Ha-ras gene.* Cancer Lett, 2000. **153**(1-2): p. 199-209.
37. Dutertre, S., et al., *Cell cycle regulation of the endogenous wild type Bloom's syndrome DNA helicase.* Oncogene, 2000. **19**(23): p. 2731-8.
38. Mohaghegh, P., et al., *The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases.* Nucleic Acids Res, 2001. **29**(13): p. 2843-9.
39. Karow, J.K., et al., *The Bloom's syndrome gene product promotes branch migration of holliday junctions.* Proc Natl Acad Sci U S A, 2000. **97**(12): p. 6504-8.
40. van Brabant, A.J., et al., *Binding and melting of D-loops by the Bloom syndrome helicase.* Biochemistry, 2000. **39**(47): p. 14617-25.
41. Johnson, F.B., et al., *Association of the Bloom syndrome protein with topoisomerase IIIalpha in somatic and meiotic cells.* Cancer Res, 2000. **60**(5): p. 1162-7.
42. Wu, L., et al., *The Bloom's syndrome gene product interacts with topoisomerase III.* J Biol Chem, 2000. **275**(13): p. 9636-44.
43. Wu, L., et al., *BLAP75/RMI1 promotes the BLM-dependent dissolution of homologous recombination intermediates.* Proc Natl Acad Sci U S A, 2006. **103**(11): p. 4068-73.
44. Wu, L. and I.D. Hickson, *The Bloom's syndrome helicase suppresses crossing over during homologous recombination.* Nature, 2003. **426**(6968): p. 870-4.
45. Plank, J.L., J. Wu, and T.S. Hsieh, *Topoisomerase IIIalpha and Bloom's helicase can resolve a mobile double Holliday junction substrate through convergent branch migration.* Proc Natl Acad Sci U S A, 2006. **103**(30): p. 11118-23.
46. Raynard, S., W. Bussen, and P. Sung, *A double Holliday junction dissolvasome comprising BLM, topoisomerase IIIalpha, and BLAP75.* J Biol Chem, 2006. **281**(20): p. 13861-4.
47. Mankouri, H.W. and I.D. Hickson, *The RecQ helicase-topoisomerase III-Rmi1 complex: a DNA structure-specific 'dissolvasome'?* Trends Biochem Sci, 2007. **32**(12): p. 538-46.
48. Chen, C.F. and S.J. Brill, *Binding and activation of DNA topoisomerase III by the Rmi1 subunit.* J Biol Chem, 2007. **282**(39): p. 28971-9.
49. Raynard, S., et al., *Functional role of BLAP75 in BLM-topoisomerase IIIalpha-dependent holliday junction processing.* J Biol Chem, 2008. **283**(23): p. 15701-8.
50. Xu, D., et al., *RMI, a new OB-fold complex essential for Bloom syndrome protein to maintain genome stability.* Genes Dev, 2008. **22**(20): p. 2843-55.
51. Singh, T.R., et al., *BLAP18/RMI2, a novel OB-fold-containing protein, is an essential component of the Bloom helicase-double Holliday junction dissolvasome.* Genes Dev, 2008. **22**(20): p. 2856-68.
52. Stavropoulos, D.J., et al., *The Bloom syndrome helicase BLM interacts with TRF2 in ALT cells and promotes telomeric DNA synthesis.* Hum Mol Genet, 2002. **11**(25): p. 3135-44.
53. Opresko, P.L., et al., *The Werner syndrome helicase and exonuclease cooperate to resolve telomeric D loops in a manner regulated by TRF1 and TRF2.* Mol Cell, 2004. **14**(6): p. 763-74.

54. Lillard-Wetherell, K., et al., *Association and regulation of the BLM helicase by the telomere proteins TRF1 and TRF2.* Hum Mol Genet, 2004. **13**(17): p. 1919-32.
55. Dejardin, J. and R.E. Kingston, *Purification of proteins associated with specific genomic Loci.* Cell, 2009. **136**(1): p. 175-86.
56. Bochkarev, A., et al., *The crystal structure of the complex of replication protein A subunits RPA32 and RPA14 reveals a mechanism for single-stranded DNA binding.* EMBO J, 1999. **18**(16): p. 4498-504.
57. Kerr, I.D., et al., *Insights into ssDNA recognition by the OB fold from a structural and thermodynamic study of Sulfolobus SSB protein.* EMBO J, 2003. **22**(11): p. 2561-70.
58. Bochkarev, A. and E. Bochkareva, *From RPA to BRCA2: lessons from single-stranded DNA binding by the OB-fold.* Curr Opin Struct Biol, 2004. **14**(1): p. 36-42.
59. Brill, S.J. and S. Bastin-Shanower, *Identification and characterization of the fourth single-stranded-DNA binding domain of replication protein A.* Mol Cell Biol, 1998. **18**(12): p. 7225-34.
60. Bochkareva, E., S. Korolev, and A. Bochkarev, *The role for zinc in replication protein A.* J Biol Chem, 2000. **275**(35): p. 27332-8.
61. Opresko, P.L., et al., *Telomere-binding protein TRF2 binds to and stimulates the Werner and Bloom syndrome helicases.* J Biol Chem, 2002. **277**(43): p. 41110-9.
62. Bachrati, C.Z. and I.D. Hickson, *RecQ helicases: suppressors of tumorigenesis and premature aging.* Biochem J, 2003. **374**(Pt 3): p. 577-606.
63. Hu, P., et al., *Evidence for BLM and Topoisomerase IIIalpha interaction in genomic stability.* Hum Mol Genet, 2001. **10**(12): p. 1287-98.
64. Wang, F., et al., *The POT1-TPP1 telomere complex is a telomerase processivity factor.* Nature, 2007. **445**(7127): p. 506-10.
65. Brunger, A.T., et al., *Crystallography & NMR system: A new software suite for macromolecular structure determination.* Acta Crystallogr D Biol Crystallogr, 1998. **54**(Pt 5): p. 905-21.
66. Jones, T.A., et al., *Improved methods for building protein models in electron density maps and the location of errors in these models.* Acta Crystallogr A, 1991. **47 (Pt 2)**: p. 110-9.
67. Bochkareva, E., et al., *Structure of the RPA trimerization core and its role in the multistep DNA-binding mechanism of RPA.* EMBO J, 2002. **21**(7): p. 1855-63.

Chapter 5

Conclusions and perspectives

Being synthesized and extended by telomerase, telomeres make sure chromosomes can be completely replicated during cell divisions. Functionally, telomeres keep the genome integrity and protect chromosome ends from degradation and being processed by inappropriate DNA repair pathways (reviewed in [1-3]). Telomeric proteins, the permanent residents at telomere region, play an important role in regulating telomerase activity and protecting telomere (reviewed in [2-4]). In my studies, my aim is to reveal the biological functions of the telomere-associated proteins by structural and biochemical analyses. So far, I have performed structural studies on three different telomere-associated protein complexes: human POT1-TPP1 complex, fission yeast Taz1 and Taz1-Rap1 complex, and human RMI1-RMI2 complex in the BLM-TOPOIIIα-RMI1-RMI2 (BTR) complex.

5.1 POT1-TPP1 complex on 3' G-overhang

By solving the crystal structure of the N-terminal portion of human TPP1, I found out that TPP1 contains one OB-fold that is structurally conserved with that in N-terminus of *Oxytricha nova* TEBP-β subunit (see Chapter 2). Since human POT1, TPP1's binding partner, was previously identified as the homologue of *Oxytricha nova* TEBP α-subunit

[5, 6], I concluded that human POT1-TPP1 complex is structurally conserved with *Oxytricha nova* TEBP α/β complex. However, beyond our anticipation, unlike the TEBP α/β complex that has a negative regulatory effect on telomerase extension [7, 8], human POT1-TPP1 complex showed a dramatic stimulation effect on telomerase activity as well as processivity in the *in vitro* telomerase activity assay (see Chapter 2). Therefore, a new question arises: what is the mechanism for human POT1-TPP1 up-regulates telomerase activity and processivity?

Previous studies already showed that POT1 could negatively and positively regulate telomere extension [9]. Generally, POT1 tends to bind to the 3'-terminal sequence GGTTAG and functions as a strong telomerase inhibitor. However, if POT1 was fixed to the 5' sequence (by mutation on the telomeric sequence) and released the unbound 3' G-overhang for more than eight nucleotides, the POT1-ssDNA complex can be extended by telomerase [9]. It seems that if POT1 is forced to stay away from the 3' terminus, it will no longer repress telomerase activity. Thus, we speculated that when TPP1 bound to POT1, POT1's position on the ssDNA would be changed. Unfortunately, our hypothesis was proved to be wrong by the SVPI enzyme digestion results: even when TPP1 was presented, majority of the POT1 molecules still located at the 3' end of ssDNA primer. This result also matched the kinetic study result that POT1-TPP1 complex would prefer to bind to the 3' end of ssDNA primer with higher affinity (K_d = 0.7 nM, see Chapter 2). However, it is still possible that there are a few POT1-TPP1 complex bind to the internal site on the primer (K_d = 7.4 nM, see Chapter 2) and leave the 3' G-overhang unoccupied for the processive telomerase extension.

Another possible solution is the telomerase recruitment model. I and other group have already detected the direct *in vitro* interaction between TPP1 and hTERT [10, 11], the protein component of telomerase. By collaborating with Dr. Neal Lue's lab in Weill Medical College of Cornell University, we also provided evidence showing that the *Candida albicans* telomerase protein Est3, the potential homologue of TPP1, may mediate a conserved function as TPP1 in telomerase recruitment and regulation [12]. Therefore, I propose that through direct interaction with TPP1, telomerase is recruited to its ssDNA substrate for telomere elongation. The direct TPP1-hTERT interaction also prevents telomerase from falling off from the extended ssDNA in the translocation step (see Chapter 1), therefore results in the increased telomerase processivity. In order to test my hypothesis, the interaction domains as well as the essential residues of TPP1 and hTERT need to be characterized first. And then the TPP1 truncations or mutants that fail to interact with telomerase will be used in the *in vitro* telomerase activity assay to check whether ssDNA primer can still be extended by telomerase with increased processivity and activity.

It is also necessary to check how POT1-TPP1 complex is organized on single-stranded telomeric DNA. As telomeric DNA being extended by telomerase, more POT1 binding sites will be generated so more POT1-TPP1 complexes will be loaded on the ssDNA. Whether the increased telomerase activity and processivity is brought by the accumulation effect of more telomerase recruitments by more POT1-TPP1 complexes? There are data suggesting that only one upstream binding POT1-TPP1 complex is enough to increase telomerase processivity (Latrick and Cech, unpublished). Furthermore, a single amino-acid mutation on the surface of TEN (telomerase essential N-terminal)

domain of hTERT would eliminate the enhanced processivity, which further suggested the specific interaction between TPP1 and telomerase (Latrick and Cech, unpublished). Besides, the structural view of how POT1-TPP1 complex localizes on longer single-stranded telomeric DNA will definitely help address the question. Currently, colleagues in our lab are working on determining the crystal structure of POT1-TPP1-ssDNA (10mer) by using the fusion protein strategy. Moreover, our preliminary electronic microscopy (EM) data and 3D reconstruction model suggested that when POT1 bound to 132mer ssDNA that contains 11 tandem POT1 binding sites, the POT1-ssDNA complex presented a cylinder-like higher-ordered structure, which further supported the notion that POT1 blocks telomerase's access to ssDNA substrate. Similarly, EM studies can be applied to exam whether the structure of the POT1-TPP1-ssDNA ternary complex is in a favorable state for telomerase extension.

5.2 Taz1, the functional homologue of human TRF1 and TRF2 in fission yeast

The double-stranded telomeric DNA binding proteins are found in budding yeast, fission yeast and mammals [13-21]. Structurally, they all share the C-terminal Myb-domain to mediate duplex telomeric DNA interaction in a sequence specific manner. Besides, the TRFH (TRF homology) domain is found in human double-stranded telomeric proteins, TRF1 and TRF2 [18]. Structural evidence reveals that the helical-TRFH domain is responsible for the homodimerization for both TRF1 and TRF2 [18]. Taz1, the only double-stranded telomeric DNA binding protein in fission yeast, is proposed to be the structural homologue of human TRF1 and TRF2, and therefore to bear the similar TRFH-like domain for its dimerization. By solving the crystal structure of $Taz1_{TRFH}$, I found out

that although $Taz1_{TRFH}$ is α-helix-rich as TRFH domains in TRF1 and TRF2, it is not responsible for Taz1 homodimerization due to the different structural architecture. Instead, Taz1 dimerizes via the formation of a four-helix-bundle through the dimerization domain that is outside $Taz1_{TRFH}$. Despite the structural differences, Taz1 still functions similarly as TRF1 and TRF2. First, Taz1 dimerizes in order to bind duplex telomeric DNA tightly. Like TRF1 and TRF2, disruption of dimerization will sweep Taz1 from the duplex telomeric DNA, as confirmed by *in vitro* binding data and *in vivo* evidence. Second, our solution structure of Taz1-Rap1, as well as the *in vivo* biochemical data, clearly show that Taz1 and Rap1 mimic their homologues TRF2 and Rap1 in human. Collectively, I concluded that Taz1 is not structurally conserved with TRF1 and TRF2 as we expected before, but it is indeed a functional homologue of TRF1 and TRF2.

We are surprised to discover the different structural architecture of Taz1 and are curious about the reason that why the homology proteins would bear such different domain structures. According to my current results, it is not accurate to take the TRFH domain as the signature motif of all double-stranded telomeric DNA binding proteins since the structure of $Taz1_{TRFH}$ is so different from those in human TRF1 and TRF2. At this stage, I cannot propose a proper explanation for the structural difference between Taz1 and human TRF1 and TRF2. It is possible that different structural architecture is developed for diverse functions. In higher organisms, during the evolution, more complicated functions are needed and might be designated to the TRF proteins. For example, higher species such as human have developed two TRF proteins: TRF1 and TRF2. Different structural architectures with distinctive functions have been rearranged in order to tolerate the evolutionary stress. Therefore, more structural studies on the

dsDNA telomeric proteins in other species (such as the ancient organism *Trypanosoma*) are required in order to address the question that how double-stranded telomeric proteins evolved during the evolution.

5.3 Long-term goals

In human, shelterin complex is made up of six telomeric proteins: TRF1, TRF2, TIN2, Rap1, POT1 and TPP1 [2]. Almost every component has its corresponding homologue in other species. For example, POT1-TPP1 is structurally conserved with TEBP α/β complex in ciliates (see Chapter 2). Rap1 is found in budding yeast, fission yeast and mammals (see Chapter 3). Taz1 is the functional homologue of TRF1 and TRF2 (see Chapter 3). Both TRF-like protein and POT1-like protein have been found in trypanosome *T. brucei* (personal communications with Dr. Li Bibo). Recently, the components of a shelterin-like complex in *S. pombe* have been identified successively. They are Tpz1, Poz1 and Ccq1 [22] (see Fig. 1.9 in Chapter 1). They form a telomere-associated complex with previous identified Taz1, Rap1 and Pot1, further suggesting that the telomere-end-protection mechanism by a telomeric-protein complex is conserved. Although the detailed biological functions of the telomeric-protein complexes in different species are still under investigation, the overall function remains the same: to protect telomere and regulate telomerase activity. To better understand telomere and telomerase, more structural and functional information of every component of shelterin-like complex and related pathways need to be characterized. So far, our lab has solved the crystal structures of most components of human shelterin complex. Meanwhile, I am exploring how telomere is maintained and processed in telomerase-negative ALT (alternative

lengthening of telomeres) cells on the structural basis. The structural study on the double Holliday junction (dHJ) dissovalsome BLM-TOPOIIIα-RMI1-RMI2 (BTR) complex, which potentially associated with telomere in ALT cells, is performed in order to help us understand how telomere is processed and metabolized under abnormal conditions (such as lose telomere end protection) via homologous recombination. The long-term goal is that by modeling all the available structures, we would be able to draw the overall structural picture showing how shelterin complex and other telomere-associated proteins protect telomeres and regulate telomerase, which will provide insights for better understand the mysterious chromosome ends.

Reference:

1. Cech, T.R., *Beginning to understand the end of the chromosome.* Cell, 2004. **116**(2): p. 273-9.
2. de Lange, T., *Shelterin: the protein complex that shapes and safeguards human telomeres.* Genes Dev, 2005. **19**(18): p. 2100-10.
3. Palm, W. and T. de Lange, *How shelterin protects mammalian telomeres.* Annu Rev Genet, 2008. **42**: p. 301-34.
4. Smogorzewska, A. and T. de Lange, *Regulation of telomerase by telomeric proteins.* Annu Rev Biochem, 2004. **73**: p. 177-208.
5. Baumann, P. and T.R. Cech, *Pot1, the putative telomere end-binding protein in fission yeast and humans.* Science, 2001. **292**(5519): p. 1171-5.
6. Baumann, P., E. Podell, and T.R. Cech, *Human Pot1 (protection of telomeres) protein: cytolocalization, gene structure, and alternative splicing.* Mol Cell Biol, 2002. **22**(22): p. 8079-87.
7. Gottschling, D.E. and V.A. Zakian, *Telomere proteins: specific recognition and protection of the natural termini of Oxytricha macronuclear DNA.* Cell, 1986. **47**(2): p. 195-205.
8. Price, C.M. and T.R. Cech, *Telomeric DNA-protein interactions of Oxytricha macronuclear DNA.* Genes Dev, 1987. **1**(8): p. 783-93.
9. Lei, M., et al., *Switching human telomerase on and off with hPOT1 protein in vitro.* J Biol Chem, 2005. **280**(21): p. 20449-56.
10. Wang, F., et al., *The POT1-TPP1 telomere complex is a telomerase processivity factor.* Nature, 2007. **445**(7127): p. 506-10.
11. Xin, H., et al., *TPP1 is a homologue of ciliate TEBP-beta and interacts with POT1 to recruit telomerase.* Nature, 2007. **445**(7127): p. 559-62.
12. Yu, E.Y., et al., *A proposed OB-fold with a protein-interaction surface in Candida albicans telomerase protein Est3.* Nat Struct Mol Biol, 2008. **15**(9): p. 985-9.
13. Conrad, M.N., et al., *RAP1 protein interacts with yeast telomeres in vivo: overproduction alters telomere structure and decreases chromosome stability.* Cell, 1990. **63**(4): p. 739-50.
14. Lustig, A.J., S. Kurtz, and D. Shore, *Involvement of the silencer and UAS binding protein RAP1 in regulation of telomere length.* Science, 1990. **250**(4980): p. 549-53.
15. Buck, S.W. and D. Shore, *Action of a RAP1 carboxy-terminal silencing domain reveals an underlying competition between HMR and telomeres in yeast.* Genes Dev, 1995. **9**(3): p. 370-84.
16. Cooper, J.P., et al., *Regulation of telomere length and function by a Myb-domain protein in fission yeast.* Nature, 1997. **385**(6618): p. 744-7.
17. Li, B., S. Oestreich, and T. de Lange, *Identification of human Rap1: implications for telomere evolution.* Cell, 2000. **101**(5): p. 471-83.
18. Fairall, L., et al., *Structure of the TRFH dimerization domain of the human telomeric proteins TRF1 and TRF2.* Mol Cell, 2001. **8**(2): p. 351-61.
19. Zhong, Z., et al., *A mammalian factor that binds telomeric TTAGGG repeats in vitro.* Mol Cell Biol, 1992. **12**(11): p. 4834-43.

20. Bilaud, T., et al., *Telomeric localization of TRF2, a novel human telobox protein.* Nat Genet, 1997. **17**(2): p. 236-9.
21. Broccoli, D., et al., *Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2.* Nat Genet, 1997. **17**(2): p. 231-5.
22. Miyoshi, T., et al., *Fission yeast Pot1-Tpp1 protects telomeres and regulates telomere length.* Science, 2008. **320**(5881): p. 1341-4.